THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1996, *49A* (3), 828–844

A Probabilistic Interpretation of "If–Then"

In-mao Liu
National Chung-Cheng University, Chia-Yi, Taiwan
King-chung Lo
Chinese University of Hong Kong, Shatin, N.T., Hong Kong
Jei-tun Wu
National Taiwan University, Taipei, Taiwan

This study proposes that subjects interpret thematic conditionals ("if *p* then *q*") probabilistically in solving conditional reasoning problems. Experiment 1 found that subjects' correct responses increased with the perceived probability of *q*, given *p* for each of the four forms of conditional arguments: modus ponens (MP), modus tollens (MT), denial of the antecedent (DA), and affirmation of the consequent (AC). Experiment 2 ruled out two alternative explanations based on the comprehensibility of conditionals and on subjects interpreting conditionals as biconditionals. In Experiment 3, subjects solved two types of problems: (a) complete probabilistic problems, such as "If *p* then *q*; knowing *p*; how probable is *q*?", and (b) reduced probabilistic problems, such as "Knowing *p*; how probable is *q*?" Two sources of information that determine the observable reasoning responses are identified. One source of information is based on one's general knowledge, and another is based on taking all premises into account.

A major finding of many experiments on conditional reasoning is that the content of premises can have striking effects on the subject's performance (e.g. Evans, Newstead, & Byrne, 1993, for review; Griggs & Cox, 1982; Johnson-Laird & Byrne, 1991, also for review; Wason & Johnson-Laird, 1972). The reason why people are much more successful at reasoning about a meaningful rule than a logically equivalent arbitrary rule is, however, not entirely clear. When arbitrarily abstract rules are used in a conditional reasoning task known as the Wason selection task, usually less than 10% of the subjects produce the correct solution (Wason & Johnson-Laird, 1972). Later studies (Cheng & Holyoak, 1985; Cosmides, 1989), however, have shown that the level of performance on the Wason selection task could be higher than 70 or 80%, if permission statements ("If the action is to be taken, then the pre-condition must be satisfied") or social-contract rules ("If you take the benefit, then you pay the cost") are used.

Requests for reprints should be sent to In-mao Liu, Department of Psychology, National Chung-Cheng University, Chia-Yi, Taiwan. FAX: 886–5–2720857; E-mail: psyiml@ccunix.ccu.edu.tw
This study was supported by the National Science Council Grant 81–0111–S–194–501. We are especially grateful to Kris Kirby and Peter H. Schoenemann for helpful comments on earlier versions of this article.

In formal logic, an indicative conditional "if *p* then *q*" means "*not-p* or *q*". This is because, as any conditional is false when *p* (its antecedent) is true and *q* (its consequent) is false, it must be that any true conditional has a false antecedent or a true consequent. In studying conditional reasoning, this truth-functional conceptualization may be useful for describing the behaviour of those sophisticated college students who have studied logic and internalized such knowledge. We consider, however, that most people take a conditional simply as a probabilistic statement.

The purpose of the present study was twofold. First, we attempted to ascertain whether subjects interpret thematic conditionals probabilistically. Second, conditional reasoning problems were rephrased into two forms of probabilistic arguments to identify two sources of information that determine observable reasoning responses. The second purpose of the present study is explained in the introduction to Experiment 3.

As for the first purpose of the present study, if conditionals are interpreted as conveying the conditional probability of *q* given *p*, the next question is how to subject them to empirical investigation. This is because "if *p* then *q*" is highly dependent on context and may merely denote the subjective confidence of *q* given *p* on pragmatic grounds (Grice, 1989). Our procedure was to select three sets of conditionals for which most people are likely to agree on different magnitudes of conditional probabilities, to go with the three sets of conditionals when they are used without any particular context. Thus, on the basis of perceived sufficiency of *p* for *q* for each statement, Liu and Lo (1990) classified conditionals into three types: high, medium, and low perceived sufficiency. These three types of conditionals may be exemplified by the following:

If John is living in Canada, then he is living in the Northern Hemisphere. (1)
If Mary catches cold, then she will take a one-day leave from the company. (2)
If Bob puts white clothes on, then he goes to the library. (3)

These three statements differ in the degree of perceived probability of *q* given *p*. In Statement 1, the fact that John is living in Canada is perceived as quite sufficient to ensure that John is living in the Northern Hemisphere, because Canada is in the Northern Hemisphere, as based on one's general knowledge. However, in Statement 2, if Mary catches cold, she may not necessarily take a one-day leave, because the cold may not be serious. Therefore, the perceived probability in this case is not high. In the case of Statement 3, the *a priori* probability of Bob going to the library if he puts white clothes on is low. Liu and Lo found that the percentages of correct MP (modus ponens) and MT (modus tollens) endorsements all increased as a function of perceived sufficiency. Experiment 1 was a replication of the Liu and Lo study with a larger group of subjects. Experiment 2 was conducted to rule out some alternative explanations for the results of Experiment 1. By rephrasing conditional arguments into probabilistic arguments, Experiment 3 estimated the percentage of subjects who interpreted conditionals probabilistically.

A probabilistic account of conditionals appeared early in philosophy (Jackson, 1987, for a review). According to Adams (1975), the assertibility of an indicative conditional is the conditional probability of its consequent, given its antecedent. Assertibility generally refers to the justifiability of saying something.

Let us consider a simple argument consisting of two premises and one conclusion, as follows:

Premises:
If it rains, then there is enough water for the flowers. (4a)
It rains. (4b)
Conclusion:
There is enough water for the flowers. (4c)

This argument is in the MP form, "If *p*, then *q*, *p*; therefore *q*". For the conditional, "if *p* then *q*", *p* is sufficient for *q* and *q* necessary for *p*. An event is either sufficient (necessary) or not sufficient (not necessary) for another event. An event cannot be partially sufficient (partially necessary) for another event in two-valued logic. However, most people may understand the first premise of the present argument as "If it rains, then probably there will be enough water for the flowers". In other words, judging from our experiences of the extent to which *p* predicts *q*, we tend to interpret "if *p* then *q*" to mean "in case *p* occurs, *q* is likely to occur to that extent". Thus, "if–then" is interpreted probabilistically, depending on the perceived sufficiency of *p* for *q*. For example, a person who lives in the Southern part of an island, where it rains heavily whenever it does rain, will consider the "if–then" in Statement 4a—the meaning of "enough water for the flowers" to follow "it rains"—very likely. On the other hand, for a person living in the Northern part, where there is only a little rain, the meaning of "if–then" in Statement 4a may be different.

In contrast to the facilitative effect of meaningful content on performance in the selection task, the probabilistic interpretation of the if–then connective attempts to explain the adverse effect of meaningful content on conditional inferences. The latter interpretation, however, is in line with Henle's (1962) view that deductive "error" is due, not to illogicality, but to premises being interpreted in an unintended way, to the introduction of outside knowledge as an additional premise, or to a failure to accept the logic task. Thus, in early experiments on conditional reasoning, abstract content was frequently used—e.g. "If the letter is A, then the number is 6" (Taplin, 1971; Taplin & Staudenmayer, 1973). The data show a consistent finding that MP is nearly perfectly endorsed. According to Henle's view, subjects are likely to accept it as a logic task if all materials are abstract. On the other hand, using meaningful content, Markovits and Savary (1992) found for the envelope problem ("If an envelope is sealed, then it must carry a 20-cent stamp") that there were only 76% correct responses in MP inferences by a group of 160 university students who served as their subjects.

EXPERIMENT 1

Experiment 1 examines the effect of perceived sufficiency on conditional reasoning. Suppose that the if–then connective is interpreted probabilistically. With the three sets of conditional statements differing in perceived sufficiency with respect to their *p*–*q* pairs, the effect of perceived sufficiency should be observed on conditional reasoning.

Method

Subjects

The subjects were 96 students, aged 15–16, at a secondary school in Taipei. They participated in the experiment voluntarily. They did not serve in any other part of the present study. They had not studied any logic course as their school subject.

Conditional Statements

Conditional statements used in the present experiment were chosen as follows: A pool of 24 conditional statements was initially created. Two judges then independently classified them into three categories (high, medium, and low), according to the perceived probability of *q* given *p*. A final list of 12 statements was determined through the following three criteria: (a) a conditional was classified by the two judges as belonging to the same category; (b) each P (p/q) was judged to be near .5; and (c) conditionals of the same category were judged to include as many as possible sorts of semantic relations holding between antecedents and consequents. The purpose of the second criterion was to equate the perceived necessity, P (p/q). With respect to the last criterion, two conditional statements involving class-inclusion relations and two involving object–property relations were selected as the "high" probability materials. They were of the same type as statements used by Ward, Byrnes, and Overton (1990) as entailment items in their study. The medium- and low-probability materials involving a wide variety of semantic relations between antecedents and consequents were relatively easy to find.

The final list consisted of the three sets of conditional statements so selected (see Appendix). They are English translations, as all conditional statements were in Chinese in the present and following experiments. Unlike the translation of a counterfactual statement that is likely to be controversial (see Au, 1983; Bloom, 1981), the translation of a simple conditional statement from Chinese into English or from English into Chinese is straightforward.

Task

Conditional reasoning problems were presented to subjects in eight argument forms (see Table 1). Such argument forms have generally been used in previous studies (e.g. Marcus & Rips, 1979). The first two argument forms were for testing MP, the second two for testing MT, the third two for testing the fallacies of DA (denial of the antecedent), and the last two for testing the fallacies of AC (affirmation of the consequent). MP and MT inferences are valid, and DA and AC are invalid.

Design and Procedure

The design was a 3×4 within-subjects factorial. The first factor was degree of perceived sufficiency (high, medium, or low), and the second factor type of argument (MP, MT, DA, or AC).

In addition to the instructions printed on the front page of a booklet distributed to each subject, the experimenter emphasized that the subjects' task was to judge whether a conclusion follows two premises for each argument. The printed instructions when translated from Chinese into English, were as follows:

> Inside this cover are 24 sets of three statements. You are asked to judge whether the third statement follows from the first and the second statements. In other words, suppose that the

TABLE 1
Argument Forms Used in Experiment 1

Type of Argument	*Argument Form*	*Validity Judgement*
MP1	If *p* then *q*; *p*. Therefore, *q*.	Valid (True)
MP2	If *p* then *q*; *p*. Therefore, *not-q*.	Invalid (False)
MT1	If *p* then *q*; *not-q*. Therefore, *not-p*.	Valid (True)
MT2	If *p* then *q*; *not-q*. Therefore, *p*.	Invalid (False)
DA1	If *p* then *q*; *not-p*. Therefore, *not-q*.	Invalid (Sometimes True)
DA2	If *p* then *q*; *not-p*. Therefore, *q*.	Invalid (Sometimes True)
AC1	If *p* then *q*; *q*. Therefore, *p*.	Invalid (Sometimes True)
AC2	If *p* then *q*; *q*. Therefore, *not-p*.	Invalid (Sometimes True)

Note: MP =Modus Ponens, MT = Modus Tollens, DA = Denial of the Antecedent, AC = Affirmation of the Consequent.

first two statements of each set are known to be true, you are to decide whether the third statement is true or not.

All sets of three statements should be worked out in the order they are presented. You are not allowed to change your previous answers. The following are three examples, the first true, the second false, and the third sometimes true and sometimes false. (Three transition inference forms followed.)

Each of the 12 conditional statements was embedded in two of the four argument types (MP, MT, DA, and AC), thereby producing 24 arguments. Assignment of each conditional to two of the four argument types was balanced between subjects. Also balanced between subjects was the assignment of each conditional to an argument form within each argument type. Each subject contributed two observations in each of 3 (degrees of perceived sufficiency) × 4 (argument types) cells. Presentation order of arguments within a booklet was randomized. When the subjects had finished reading the instructions, they were allowed to begin to work at their own pace.

Results and Discussion

The mean percentage of correct endorsements is presented in Table 2 for each type of argument for each level of perceived sufficiency, P (q/p). It is clear from Table 2 that correct responses increased with P (q/p) for each argument type. An analysis of variance showed that P (q/p) had a significant effect, $F(2, 190) = 59.38$, $p < .01$. Type of argument was also a significant source of variance, $F(3, 285) = 18.62$, $p < .01$. The interaction between P (q/p) and type of argument was also significant, $F(6, 570) = 8.34$, $p < .01$. Further analyses of the simple main effects showed that the effect of P(q/p) was signi-

TABLE 2
Percentage of Correct Endorsements as a Function of Perceived Sufficiency and Type of the Second Premise

	Type of Second Premise			
Perceived Sufficiency	*p* (MP)	*not-q* (MT)	*not-p* (DA)	*q* (AC)
High	0.94	0.84	0.76	0.73
Medium	0.86	0.80	0.52	0.65
Low	0.72	0.50	0.58	0.56

Note: MP = Modus Ponens, MT = Modus Tollens, DA = Denial of the Antecedent, AC = Affirmation of the Consequent.

ficant for MP, MT, DA, and AC, $Fs(2, 760) = 15.20$, 44.95, 21.06, and 9.62, respectively, all $ps < .01$. According to Tukey's HSD test, any two conditions of perceived sufficiency were significantly different, $p < .01$.

The effectiveness of the variable of P (q/p) replicates the Liu and Lo (1990) finding obtained from a smaller sample of subjects and extends the effect of entailment found by Ward et al. (1990), because the latter investigators were concerned with the effect of entailment only on the Wason selection task performance.

As the variable "number of counterexamples" (*p* true but *q* false) is inversely related to perceived sufficiency, the obtained effect of P (q/p) on MP and MT inferences is in line with the Cummins, Lubart, Alksnis, and Rist (1991) finding of the effect of number of counterexamples.

The theory of pragmatic reasoning schemas (Cheng & Holyoak, 1985) has difficulty in explaining the observed effect of perceived probability. The statement that a person living in a certain location is living within a larger area does not seem to represent a permission schema, nor an obligation schema. The obtained effect of the variable of perceived probability is also difficult to explain in terms of social-contract theory (Cosmides, 1989). It is hard to see how any kind of social contract is involved in a statement about a person staying arbitrarily in some place.

EXPERIMENT 2

In Experiment 1, it was found that correct reasoning increased with P (q/p) for each of the four argument types, MP, MT, DA, and AC. A question arises as to whether this finding was produced by P (q/p) alone. More specifically, the finding of Experiment 1 may be due to the comprehensibility of thematic conditionals: conditionals of high P (q/p) may be easier to comprehend than those of low P (q/p), which may have produced the obtained effect.

Furthermore, the manipulation of perceived probability in Experiment 1 was based on two judges classifying a set of conditionals into three categories (high, medium, and low). It is desirable to obtain a mean rating of perceived probability for each category of conditionals from a group of subjects. In the present experiment, groups of subjects rated P (*p*/*q*), P (*not-p*/*not-q*), and P (*not-q*/*not-p*), as well as P (*q*/*p*). As the last P (*q*/*p*) is related to MP, the former three should be related to AC, MT, and DA, respectively.

Method

Subjects and Conditional Statements

The subjects were 43 students (Grade 10—mean age about 16) at a Hong Kong high school and 40 students (Grade 10—mean age about 16) at another Hong Kong high school. They participated in the experiment voluntarily. The three sets of conditional statements listed in the Appendix were used.

Procedure

The 43 students at one school were tested in a group in a classroom. They received two practice sentences before rating the comprehensibility of each of the 12 conditional statements according to a five-point scale. The subjects read each statement as quickly as possible. At the end of reading each statement, if the subject considered it necessary to re-read a statement to ascertain its meaning, the subject rated it as "difficult to comprehend" or "very difficult to comprehend". On the other hand, if the subject after reading a statement quickly was confident of the meaning of the statement without re-reading it, the subject rated it as "easy to comprehend" or "very easy to comprehend".

After rating the comprehensibility of the 12 statements, about half the subjects (22) rated the perceived probability of *q* given *p*, P (*q*/*p*); the remaining subjects (21) rated P (*p*/*q*) for each *p*–*q* pair of the 12 statements.

Before rating P (*q*/*p*) for each *p*–*q* pair, the subjects received a practice item, "Knowing that John left home, how probable is it that he went to school?" The experimenter told the subjects that most people would consider the probability of John's going to school to be very high if John were a schoolboy. However, the same probability would be very small if John were a businessman. Then the subjects indicated their perceived probability on a six-point scale that ranged from 0.0 to 1.0. The forms of questions used in measuring all perceived probabilities are presented in Table 3.

In rating P (*q*/*p*), the subjects answered the question, "Knowing *p*, how probable is *q*?", by indicating their perceived probability on the scale for each *p*–*q* pair from a conditional statement (see Table 3). The subjects who rated P (*p*/*q*), following the same practice statement indicated their perceived probability by answering the question, "Knowing *q*, how probable is *p*?"

All the subjects rated the comprehensibility of each statement and the perceived probability about two events derived from each statement printed in a booklet. Once the subjects started to rate a new statement, they were not to go back to the statements already rated. The order of the 12 statements was randomized for each subject. There was no time limit for each subject. However, all subjects completed the rating task within half an hour.

TABLE 3
Problem Forms for Measuring Perceived Probability

Perceived Probability	*Problem Form*
P (*q*/*p*)	Knowing *p*, how probable is *q*?
P (*not-p*/*not-q*)	Knowing *not-q*, how probable is *not-p*?
P (*not-q*/*not-p*)	Knowing *not-p*, how probable is *not-q*?
P (*p*/*q*)	Knowing *q*, how probable is *p*?

The procedure of measuring P (*not-p/ not-q*) and P (*not-q/ not-p*) was essentially the same as that for measuring P (*q/ p*) and P (*p/ q*). Half the 40 subjects at the other school rated P (*not-p/ not-q*); the remaining half rated P (*not-q/ not-p*).

Results and Discussion

The results of ratings are presented in Table 4. With respect to the first column, comprehensibility, the mean ratings were converted to a range of from 0.00 to 1.00, with lower values standing for easy comprehensibility and higher values for difficult comprehensibility. It is clear from the results of the first column that subjects rated all the conditional statements as extremely easy to comprehend. An ANOVA showed that the differences among the three means were not significant, $F(2, 164) < 1$.

As far as the perceived probabilities of *q* given *p* are concerned, in accordance with perfect matches by the two judges, the differences between any two mean ratings were at least 0.25 apart. An ANOVA showed that the differences were significant, $F(2, 42) = 78.64$, $p < .001$. According to Tukey's HSD test, any two means were significantly different, all *p*'s < .01.

In considering P (*not-p/ not-q*), which is the inverse of P (*q/ p*) and logically equivalent, it can be seen from the third column of Table 4 that the variable P (*q/ p*) had a significant effect. An ANOVA showed that the effect was significant, $F(2, 38) = 16.48$, $p < .001$. Any two mean ratings were at least 0.11 apart. Tukey's HSD test showed that any two means were significantly different, $p < .05$ for the high and medium conditions and $p < .01$ for the medium and low conditions.

In the fourth column of Table 4 are presented P (*not-q/ not-p*). An ANOVA showed that an overall difference was significant, $F(2, 38) = 4.105$, $p < .05$. According to the Tukey HSD test, P (*not-q*)/ *not-p*) in the high sufficiency condition was smaller than those in the medium and low sufficiency conditions ($p < .05$).

In the fifth column are listed the mean ratings of P (*p/ q*), which is the converse of P (*q/ p*). An ANOVA showed that an overall difference among the mean ratings was not significant, $F(2, 40) = 1.71$, $p > .10$.

The results of Experiment 2 showed that some possible confoundings in Experiment 1 can be ruled out. First, comprehensibility of the conditionals was found to be not

TABLE 4
Mean Ratings for Comprehensibility and Perceived Probability

Conditional Statement	*Comprehensibility*		*Perceived Probability* P (*q/ p*)		P (*p′ / q′*)		*P (q′ /p′)*		*P (p/ q)*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High Sufficiency	.10	.13	.89	.16	.71	.21	.43	.16	.47	.13
Medium Sufficiency	.11	.17	.64	.16	.60	.09	.52	.18	.47	.14
Low Sufficiency	.12	.14	.37	.18	.46	.13	.51	.18	.43	.14

Note: The standard deviation for each condition is shown in parentheses. *p′* = *not-p*, and *q′* = *not-q*.

systematically related to the perceived probability of *q* given *p*. Therefore, the finding of Experiment 1 that correct responses increased with P (*q*/*p*) cannot be attributed to the conditionals of high P (*q*/*p*) being more comprehensible than those of low P (*q*/*p*). Second, P (*p*/*q*), the converse of P (*q*/*p*), was found to be not systematically related to P (*q*/*p*). Moreover, P (*not-q*/*not-p*) was significantly lower in the high condition than in the other conditions. These findings ruled out the possibility that the conditionals of high P (*q*/*p*) in Experiment 1 were interpreted biconditionally.

It can be particularly noted from Table 4 that, in the condition of high perceived sufficiency, P (*q*/*p*) and P (*not-p*/*not-q*) are nearly twice as large as P (*not-q*/*not-p*) and P (*p*/*q*). The reason why P (*not-q*/*not-p*) was smaller in the high sufficiency condition than in the other conditions may be due to the subjects' underestimate of this probability in the high sufficiency condition. This may be because P (*q*/*p*) was already very high in the high sufficiency condition, and the underestimate would result from a contrast effect. Thus, subjects would be as uncertain of inferring "not very hard" from "not a diamond" as "not add any furniture" from "not move". However, subjects are very certain of inferring hardness from a diamond; *in contrast*, they could be more uncertain of inferring "not hard" from "not a diamond". This contrast effect would not apply to the move-furniture case, because the move-furniture inference is not very certain in the first place.

EXPERIMENT 3

Suppose that subjects tend to interpret thematic conditionals probabilistically. Then, instead of asking subjects *whether* the conclusion follows the premises, it is more natural to ask *how probable it is* that the conclusion follows the premises. When the four forms (MP, MT, DA, and AC) of conditional reasoning are rephrased in this way, four probabilistic problem forms may be obtained. These problems will be referred to as "complete" problems, because each problem still consists of two premises and a conclusion, although the conclusion is written probabilistically (see Table 5). By deleting the first premise from a complete problem but retaining the second, a "reduced" problem is obtained (also see Table 5).

TABLE 5
Complete and Reduced Probabilistic Reasoning Problems

Complete Problem Form	*Reduced Problem Form*
MP: If *p*, then *q*. Knowing *p*, how probable is *q*?	MP: Knowing *p*, how probable is *q*?
MT: If *p*, then *q*. Knowing *not-q*, how probable is *not-p*?	MT: Knowing *not-q*, how probable is *not-p*?
DA: If *p*, then *q*. Knowing *not-p*, how probable is *not-q*?	DA: Knowing *not-p*, how probable is *not-q*?
AC: If *p*, then *q*. Knowing *q*, how probable is *p*?	AC: Knowing *q*, how probable is *p*?

Note: MP = Modus Ponens, MT = Modus Tollens, DA = Denial of the Antecedent, AC = Affirmation of the Consequent.

Consider a complete problem of the form ''If *p* then *q*; knowing *p*; how probable is *q*?'' If the subject does not interpret the conditional probabilistically, then the subject may assign *q*'s probability as one in responding to this problem. Therefore, the percentage of subjects who assign values between 0 and 1 (but not including 0 and 1) in answering a set of complete probabilistic MP problems may be taken as an index of expressing a probabilistic interpretation. The first aim of Experiment 3 was to ascertain more directly whether subjects typically interpret thematic conditionals probabilistically.

By creating the reduced and complete problems, it is now possible to isolate a knowledge-based component from a premise-based component in observable reasoning performance as follows: We first assume that subjects interpret thematic conditionals probabilistically. Then, in responding to the reduced problem, ''Knowing that John moves, how probable is it that he adds some furniture?'', suppose that a high probability rating is obtained. This high probability rating must be based on the subject's knowledge that a person is likely to add some furniture whenever he or she moves.

Suppose further that there is an upward increase in the probability rating in responding to the complete problem, ''If John moves, then he adds some furniture. Knowing that John moves, how probable is it that he adds some furniture?'' Then, this increase in probability ratings reflects a component in conditional reasoning that takes all premises into account. Therefore, by introducing the reduced and complete problems, Experiment 3 attempted to identify a knowledge-based component and a premise-based component in conditional reasoning.

Method

Subjects

The subjects were 48 students, aged 16–17, at a secondary school in Chia-Yi, a small city in the south of Taiwan. They served in the experiment voluntarily.

Tasks

Two types of tasks were used. For the first type of task, MP, MT, DA, and AC were rewritten into probabilistic forms to obtain "complete problems", as presented in the left panel of Table 5. Clearly, the probabilistic forms differ from the original forms only in adding "how probable" to the conclusion of each argument.

For the second type of task, reduced forms of the complete problems were used. In the reduced forms, the first premises (conditionals) were simply deleted, as presented in the right panel of Table 5. Thus, the problem forms used for obtaining P (*q*/*p*), P (*not-p*/*not-q*), P (*not-q*/*not-p*), and P (*p*/*q*) in Experiment 2 were the reduced probabilistic argument forms.

Procedure

Subjects were tested in large groups. They worked out two practice problems printed on the front page of a booklet before attempting to solve 48 experimental problems. The first practice problem was in the *reduced* form, "Knowing that Mary is an A High School student, how probable is it that she is going to a picnic today?" They answered the problem by indicating their judged probability on an

11-point scale that ranged from 0 to 100, with 0 standing for "absolutely improbable" and 100 for "absolutely probable". The second practice problem was in the *complete* form, "If Mary is an A High School student, then she is going to a picnic today. Knowing that Mary is an A High School student, how probable is it that she is going to a picnic today?" Then subjects solved the 48 experimental problems at their own pace.

For half the subjects, the first 24 experimental problems were in the reduced form, the last 24 in the complete form. The order was reversed for the remaining subjects. The experimental problems were constructed from the conditional statements used in Experiment 1. It should be noted that there were four *p–q* pairs of high sufficiency, four *p–q* pairs of medium sufficiency, and four *p–q* pairs of low sufficiency. Therefore, for each subject the first set of 24 experimental problems was constructed by randomly selecting two out of each set of four *p–q* pairs of different degrees of perceived sufficiency. As each *p–q* pair could be used for constructing four types of arguments (MP, MT, DA, and AC), altogether there resulted 24 experimental problems. The complementary set of six *p–q* pairs was used to construct the second set of 24 experimental problems.

Results

Probabilistic Interpretations of Conditionals. Table 6 presents the percentages of subjects giving probabilistic interpretations for various sets of conditionals. A subject was counted as giving a probabilistic interpretation for a set of conditionals if the obtained probability ratings did not consistently assume extreme values (1's in the present case) for a derived set of complete MP problems.

It is clear from Table 6 that about one third of subjects interpreted the high-sufficiency conditionals probabilistically. This is apparently an underestimate, because the same one third of subjects responded probabilistically to the reduced problems that did not include an if–then statement as a premise. A better estimate should be obtained from the medium-sufficiency and low-sufficiency conditionals. For the medium-sufficiency and low-sufficiency conditionals, 60 and 87% of subjects, respectively, gave probabilistic interpretations. The percentage of subjects who responded probabilistically to all conditionals was very large—i.e. 90%.

Reduced vs. Complete Problems. As the patterns of results obtained for the reduced and complete problems did not depend on their order of administration, kind of task (reduced vs. complete problems) was considered as a within-subjects variable in the

TABLE 6
Percentage of Subjects Giving Probabilistic Interpretations

	Type of Conditional				
Kind of Problem	*HS*	*MS*	*LS*	*HS and MS*	*HS and MS and LS*
Reduced	0.35	0.94	1.00	0.94	1.00
Complete	0.35	0.60	0.87	0.67	0.90

Note: HS = High Sufficiency, MS = Medium Sufficiency, LS = Low Sufficiency.

following analyses. Table 7 presents the mean probability judgements as a function of perceived sufficiency, kind of task, and type of argument.

Several observations can be made simply by inspecting Table 7. First, the pattern of results obtained with the reduced problems is quite similar to that obtained in Experiment 2. Thus, with respect to MP and MT, the effect of perceived sufficiency was reflected in decreasing probability judgements from high to medium to low perceived sufficiency. As for DA and AC, the probability judgements were lower in the high than in the medium and low conditions. Second, the pattern of results obtained with complete problems is similar to that obtained in Experiment 1. In this comparison, it should be noted that correct endorsements were recorded for DA and AC in Experiment 1 (Table 2), but simple probability judgements were entered for DA and AC in the present experiment. For probability judgements, the higher the value, the stronger is the tendency to commit fallacies, although probability judgements do not represent truth-functional reasoning responses. Third, probability judgements generally increased from the reduced to the complete problems, although this increase seemed negligible in the high-sufficiency condition.

ANOVAs were performed separately on the mean ratings obtained in the high-, medium-, and low-sufficiency conditions. With respect to the high condition, type of argument (MP, MT, DA, or AC) was the only significant source of variation, $F(3, 141) = 172.45$, $p < .01$. Kind of task (reduced vs. complete problems) was not significant, $F < 1$, nor was the interaction between type of argument and kind of task, $F < 1$. Planned comparisons showed that none of the increases in probability ratings from the reduced to the complete problems was significant.

With respect to the medium-sufficiency condition, an ANOVA showed that kind of task had a significant effect, $F(1, 47) = 12.77$, $p < .01$. Type of argument had a significant effect, $F(3, 141) = 45.28$, $p < .01$. Their interaction was not significant, $F(3, 141) = 2.44$, $p > .06$. As our main interest was in finding out how the probability ratings increased from the reduced to the complete problems, planned comparisons showed that this

TABLE 7
Mean Probability Ratings for the Reduced and Complete Problems

Perceived Sufficiency	*Kind of Task*	*Type of Argument* MP	MT	DA	AC
High	Reduced	.94	.87	.53	.48
	Complete	.95	.88	.58	.49
	(Increase)	.01	.01	.05	.01
Medium	Reduced	.75	.69	.71	.55
	Complete	.88	.75	.73	.66
	(Increase)	.13*	.06*	.02	.11*
Low	Reduced	.61	.57	.56	.71
	Complete	.71	.62	.64	.73
	(Increase)	.10*	.05	.08*	.02

Note: MP = Modus Ponens, MT = Modus Tollens, DA = Denial of the Antecedent, AC = Affirmation of the Consequent. "*" indicates a significant increase.

increase was significant only for the MP, MT, and AC cases: $t(47) = 4.09$, $p < .01$; $t(47) = 1.77$, $p < .05$; and $t(47) = 3.22$, $p < .01$; respectively.

As for the low-sufficiency condition, an ANOVA showed that, as in the medium-sufficiency condition, kind of task had a significant effect, $F(1, 47) = 4.32$, $p < .05$. Type of argument also had a significant effect, $F(3, 141) = 20.66$, $p < .01$. The interaction between kind of task and type of argument was not significant, $F(1, 141) = 1.35$, $p > .20$. Planned comparisons showed that the increase in probability ratings from the reduced to the complete problems was significant only for the MP and DA cases—$t(47) = 2.54$, $p < .01$; $t(47) = 1.76$, $p < .05$, respectively.

Discussion

It is now clear that subjects typically interpret thematic conditionals of medium and low sufficiency probabilistically. Perhaps people tend to use a universal affirmative such as "A diamond is very hard", instead of using the if–then connective in expressing it as a conditional statement in the high-sufficiency condition.

If adjusted for the translation from probability ratings to correct endorsements, the pattern of results obtained in Table 2 is remarkably similar to that obtained from the complete probabilistic forms in Table 7. We are now in a position to explain how an addition of the premise (conditional statement) leads subjects to revise their probabilities. In the following, let us compare the results obtained from the reduced and the complete problems with respect to different conditions of perceived sufficiency.

High Perceived Sufficiency. When the premise, "if *p* then *q*", is interpreted probabilistically, it can be combined with the reduced argument forms to produce MP, MT, DA, and AC arguments in probabilistic forms. In the high-sufficiency condition, the present results showed that subjects responded to the complete problems as they did to the reduced problems. For instance, subjects responded to "Knowing that a substance is a diamond, how possible is it that it is hard?" as they did to its corresponding complete problem. This makes sense, because in the former the premise ("If a substance is a diamond, then it is hard") is a sort of internalized world knowledge, and it is not necessary to remind subjects of it in the form of a premise.

The same interpretation applies to the case of probabilistic MT problems. To take the same example, in responding to "Knowing that a substance is not hard, how probable is it that it is not a diamond?" the subject's estimated probability would be large, because everyone knows that "if it is a diamond, then it is hard". Similarly, in responding to "Knowing that a substance is not a diamond, how probable is it that it is not hard?" and "Knowing that a substance is hard, how probable is it that it is a diamond?", the subject's ratings would be near .50 or smaller, because everyone knows that even if a substance is hard, its probability of being a diamond is not large.

Medium and Low Perceived Sufficiency. In the medium and low conditions, the mean ratings obtained for the reduced problems are also comparable to those obtained in Experiment 2, and the mean ratings obtained for the complete problems are comparable to those obtained in Experiment 1 when the different measures used are taken into account.

The increase in probability ratings from the reduced to the complete problems was significant for the MP case in the medium and low conditions. Thus, in responding to "Knowing that John moves, how probable is it that he adds some furniture?", the rating may be high but not very high, because the new house may be furnished, etc. Under this circumstance, given the premise "If John moves, then he adds some furniture", subjects may take connective "if–then" to mean that "If John moves, he is *highly likely* to add some furniture". This upward probability adjustment could have occurred in the high sufficiency condition. It was obscured simply because through world knowledge the probability of *q* given *p* is as high as after supplying additional information of "if *p*, then *q*".

In the MT case, the increase in probability ratings from reduced to complete problems was significant for the medium condition but not significant for the low condition. In responding to "Knowing that John does not add any furniture, how probable is it that he does not move?", the mean rating was high (.69). This rating increased to .75 when the premise, "If John moves, then he adds some furniture", was added. This increase in the mean rating was significant, but it was small and not significant when *p* and *q* were not sufficiently related, as in the low condition.

The finding that the probabilistic measures of DA and AC fallacies were less apparent in the high condition but tended to manifest themselves in one way or another in the medium and low conditions may be explained by the contrast effect proposed earlier.

GENERAL DISCUSSION

By identifying a conditional as expressing the probability of *q* given *p*, the present view complements the model-theoretic (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) and formal (e.g. Braine, 1978; Braine & O'Brien, 1991; Macnamara, 1986; Osherson, 1975; Rips, 1983) views that consider conditionals as basically truth-functional. Although the concept of counterexamples as well as the idea of taking conditionals to mean "*not–p* or *q*" may be useful for studying the reasoning performance of those who have received logic training, as Over (1993) remarked, in real life an indicative conditional tends to be asserted and accepted when its consequent seems highly probable, given its antecedent.

In solving thematic conditional reasoning problems, a body of evidence from the present experiments indicates that subjects typically interpret conditionals probabilistically. Two sources of information are identified to produce the patterns of responses obtained from the reduced and complete problems. First, in solving a reduced problem, the subject estimates the probability of the conclusion on the basis of a single premise (the second of the two premises in the original reasoning problem). The estimated probability reflects a knowledge-based component. Second, the estimated probability is revised upward by taking another premise (the conditional statement) into account in solving a complete problem that contains two premises, as in the original reasoning problem. This upward increase in the probability, if there is any, must reflect a premise-based component, because it is obtained by taking all premises into account.

On the basis of the magnitude of the revised probability estimate, the subject comes to decide whether the conclusion of a conditional argument is to be endorsed.

Knowledge-based Component

It was observed in Experiment 1 that the percentage of correct reasoning generally increased with perceived sufficiency. Judging from the pattern of probability estimates obtained in solving the reduced problems, it may be concluded that a major component of the effects in Experiment 1 obtained by manipulating the variable of perceived sufficiency is due to the knowledge-based component.

In perceiving event *q* to follow event *p* with high probability, subjects tend to reason that *q* follows *p* with high probability. This is perhaps the fundamental principle that accounts for the obtained effect of perceived sufficiency in Experiment 1. As in the case of MP, the same principle may apply to explain the high performance level for MT in the high condition, because the probability of *not-p* following *not-q* was estimated to be quite high. A question arises as to how people learn to perceive that the probability of *not-p* following *not-q* is high when the probability of *q* following *p* is high.

We propose the following account of how people learn MT inferences that are affected by perceived sufficiency. Let us consider the conditional, "If it is a diamond, then it is hard". The argument proceeds as follows:

A diamond is a hard thing. (5a)
A diamond is not a soft (or not-hard) thing. (5b)
A soft (not-hard) thing is not a diamond. (5c)
Therefore, if it is not hard, then it is not a diamond. (5d)

The crucial argument is from Statement 5b to Statement 5c. This argument is valid, because the converse of a universal negative is true if the negative is true. In other words, if "B is not C", then "C is not B".

Suppose that "B is not C" is not true sometimes. It is then not always the case that "C is not B". This explains why MT inferences are also affected by perceived sufficiency. For example, we know that, whenever a person moves, he or she tends to add some furniture. In this case, by observing that a person did not add any furniture, we conclude that it is not very likely that he or she moved.

Premise-based Component

The evidence for this component is based on the finding that subjects adjusted the probability estimates upward in solving the complete problems when an additional premise (conditional statement) was introduced. Although no such upward increase in the probability estimates was observed in the high condition, it must be due to a ceiling effect, because the probability estimates obtained in solving the reduced MP and MT problems were already quite high. The reason why the premise-based component was so small for MT inferences in the low condition may be because *q* was not obviously related to *p*. As

Adams (1975) hypothesized, "if *p* then *q*" is assertible if and only if the probability of *q* given *p* is high.

As for DA and AC inferences, because the *p–q* relationship expressed in the conditional statements in the high condition is already part of subjects' world knowledge, their introduction should not cause any increase in the probability estimates. This is what was found in Experiment 3. In the medium and low sufficiency conditions, on the other hand, there was a room for adjusting the probability estimates for the MP and MT cases. When this tendency is generalized to the DA and AC cases, an upward increase in the probability estimates results in more reasoning errors by translating probability into endorsement measures.

REFERENCES

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

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

Bloom, A.H. (1981). *The linguistic shaping of thought: A study in the impact of language in China and the West*. Hillsdale NJ: Lawrence Erlbaum Associates, Inc.

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

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

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

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187–276.

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

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

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Griggs, R.A., & Cox, J.R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73*, 407–420.

Henle, M. (1962). The relation between logic and thinking. *Psychological Review, 69*, 366–378.

Jackson, F. (1987). *Conditionals*. Oxford: Blackwell.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hove, U.K.: Lawrence Erlbaum Associates Ltd.

Liu, I.M., & Lo, K.C. (1990). *A perceived-sufficiency account of conditional reasoning*. A paper presented at the Annual Meeting of the Psychonomic Society.

Macnamara, J. (1986). *A border dispute: The place of logic in psychology*. Cambridge, MA: M.I.T. Press.

Marcus, S.L.., & Rips, L.J. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behaviour, 18*, 199–233.

Markovits, H., & Savary, F. (1992). Pragmatic schemas and the selection task: To reason or not to reason. *Quarterly Journal of Experimental Psychology, 45A*, 133–148.

Osherson, D. (1975). Logic and models of logical thinking. In R.J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Over, D.E. (1993). Deduction and degrees of belief. *Behavioral and Brain Sciences, 16*, 361–362.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*, 38–71.

Taplin, J.E. (1971). Reasoning with conditional sentences. *Journal of Verbal Learning and Verbal Behavior, 10*, 219–225.

Taplin, J.E., & Staudenmayer, H. (1973). Interpretation of abstract conditional sentences in deductive reasoning. *Journal of Verbal Learning and Verbal Behavior, 12*, 530–542.
Ward, S.L., Byrnes, J.P., & Overton, W.F. (1990). Organization of knowledge and conditional reasoning. *Journal of Educational Psychology, 82*, 832–837.
Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Original manuscript received 19 July 1994
Accepted revision received 17 August 1995

APPENDIX

Conditional Statements

High Perceived Sufficiency

If H1 lives in Canada, then HI lives in the Northern Hemisphere.
If a substance is a diamond, then the substance is very hard.
If H2 is five years old, then H2 is a child.
If H3 is a nurse, then H3 is a member of the medical personnel.

Medium Perceived Sufficiency

If M1 moves, then M1 adds some furniture.
If M2 comes back home late, then M2 will be scolded by his wife.
If M3 catches cold, then M3 will take one-day leave from the company.
If M4 cheats in the exam, then M4 will be punished by the school.

Low Perceived Sufficiency

If L1 wears glasses, then L1 is intelligent.
If L2 is at home, then L2 watches a TV.
If L3 puts white clothes on, then L3 goes to the library.
If L4 is hungry, then L4 thinks about a hamburger

Note: H1, M2, or L3 stands for a boy's or girl's name in Chinese.