Processing of Logically Valid and Logically Invalid Conditional Inferences in Discourse Comprehension

Aaron W. Rader and Vladimir M. Sloutsky
The Ohio State University

Two competing theories of processing of conditionals (if-then) were tested. Syntactic theories posit that people only draw inferences conforming to the logically valid modus ponens (MP) schema. Mental models theories predict that people draw MP and invalid affirming-the-consequent (AC) inferences. Three experiments tested these predictions. Participants read short stories that conformed to either the MP or AC form but without conclusions, and they completed either priming or recognition tasks. Results indicate that both MP and AC inferences occur during discourse processing: MP and AC premise forms prime their respective conclusions, participants erroneously judged that they had read the conclusions to MP and AC arguments, and AC inferences did not stem from a biconditional interpretation of conditionals. Findings support mental models theories.

The logical connective if . . . then, often used to express conditional statements, is ubiquitous in human discourse. Its role is particularly prominent in comprehension and reasoning. If . . . then is hypothesized to play an important role in language comprehension by enhancing the integration of discourse into a coherent representation (e.g., Braine, 1990; Halliday & Hasan, 1976) and by expressing causal, hypothetical, and pragmatic relationships (e.g., Fillenbaum, 1986; Halliday & Hasan, 1976; Sanford, 1989). If . . . then also plays a critical role in reasoning, allowing people to derive conclusions from a combination of conditional (e.g., If A then B) and categorical premises (e.g., A). However, the manner in which conditionals are processed remains a subject of debate. The goal of the present research was to shed light on this processing.

In standard propositional logic, conditional arguments may yield both warranted and unwarranted conclusions. Those forms yielding warranted conclusions are valid, whereas those yielding unwarranted conclusions are invalid. Examples of such argument forms are displayed in Table 1. As the table illustrates, the inference forms of modus ponens (MP) and modus tollens (MT) are logically valid. On the other hand, arguments of the form affirming the consequent (AC) and denying the antecedent (DA) are logically invalid, and inferences should not be drawn from these arguments.

The theoretical debate highlights two possibilities for how people process discourse containing conditional statements. One possibility is that people extract the syntactic structure of a conditional argument, and whenever this structure matches a stored inference schema, the person infers the conclusion automatically. This case, reasoning competence must express itself directly in comprehension. Another possibility is that people extract a semantic structure from a conditional argument by representing states of affairs, or models, compatible with the argument. Because these models may not be logically veridical, logically unwarranted inferences may occur during comprehension. Therefore, the analysis of conditional processing during comprehension offers a way to better understand reasoning and its mechanism.

Syntactic Theories

The syntactic, or mental logic, approach to the processing of conditional arguments holds that there is a largely innate logical competence underlying people’s ability to reason logically and to have intuitions of logical necessity (e.g., Braine, 1978, 1990; Fodor, 1975; Macnamara, 1986; Rips, 1994; see also Stein, 1996). Several syntactic theories propose that logically untrained individuals possess procedural inference schemata, stored in the lexical entries for logical connectives, that apply to the propositional logical form of incoming discourse (Braine & O’Brien, 1991; Lea, 1995; Lea, O’Brien, Fisch, Braine, & Noveck, 1990; see also Rips, 1995). These schemata are procedural because they should apply “spontaneously, without goals, goal setting, or reasoning strategies” (Lea et al., 1990, p. 363). Not all schemata proposed by syntactic theorists should apply automatically to discourse, so the following discussion pertains only to those that do.

The basic inference process according to syntactic theories is as follows. A person first automatically abstracts the propositional logical form of the discourse on the basis of activation of inference schemata stored in the lexical entries for logical connectives such as if . . . then. One inference schema (Braine & O’Brien, 1991;
Someone processing that story should infer that it was night even without instruction to draw inferences. Thus, reasoning competence should manifest itself during the process of ordinary discourse comprehension.

Several studies have provided evidence suggesting that people do draw MP inferences during comprehension. Participants given stories like those in Table 2 falsely but reliably recognized such inferences as having been presented in the stories (Franks, 1996, 1997; Lea et al., 1990). In addition, results of priming experiments with a lexical decision task have indicated that participants draw MP inferences as having been presented in the stories (Franks, 1996, 1997; Lea, 1995; Rips, 1994) is that for MP. Whenever input matches the propositional form of this schema (If then ; P) the conclusion (Q) should be drawn automatically. For example, Sentence 2 and the inference sentence in the story in the top third of Table 2 contain propositions conforming to the MP inference schema. Someone processing that story should infer that it was night even without instruction to draw inferences. Thus, reasoning competence should manifest itself during the process of ordinary discourse comprehension.

A difficulty for syntactic theories is that many studies have shown that participants drew unwarranted conditional conclusions such as AC and DA (see Evans, Newstead, & Byrne, 1993). To accommodate these findings, syntactic theories make additional assumptions. One proposal (Braine & O’Brien, 1991; Rumain, Connell, & Braine, 1983) is that invalid inferences arise through other, pragmatic invited inferences (Geis & Zwicky, 1971; Grice, 1975) that a person makes. According to this hypothesis, a conditional of the form If then leads one to infer that If not then . Thus, hearing “If you mow the lawn, I’ll give you $5” leads one to believe that “If you don’t mow the lawn, I won’t give you $5” is also true (Geis & Zwicky, 1971). Under such an interpretation, both valid (e.g., MP) and invalid (e.g., DA) inferences seem warranted.

A second syntactic hypothesis posits that a conditional may be misrepresented as a biconditional (if and only if). Specifically, people may maintain two variants of the conditional premise: (a) If then Q and (b) If then (note that syntactic theories would make this hypothesis because these theories do not include procedural inference schemata corresponding directly to the biconditional form). In this case, either categorical premise (i.e., P or Q) would supply premises conforming to the MP form, thus licensing a valid inference. Thus, an AC inference may instead be an MP inference based on Representation b (If then P) of the conditional premise. Both hypotheses are largely untested; however; invited inferences have not been demonstrated to occur online during comprehension, and no clear evidence exists to support the biconditional interpretation.

Therefore, because syntactic theories lack a procedural inference schema for either the biconditional or any invalid inference, the occurrence of invalid inferences, such as AC, during comprehension would be difficult to capture within a purely syntactic approach. Syntactic theories thus seem to predict that (a) reasoning is largely a component of comprehension processes; (b) MP inferences should occur during comprehension, whereas AC inferences should not occur or should occur much less frequently than MP inferences; and (c) whenever AC inferences do occur, they should reflect the biconditional interpretation of conditionals.

### Semantic Theories

The second approach to conditional inference is semantic in essence and is best exemplified by the mental models theory of inference (Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeeken, 1992). According to this theory, people do not parse incoming discourse into propositional logical syntax; hence, inferences drawn are not based on procedural schemata that correspond to formal logic. Instead, mental models theory proposes a three-step process of inference. First, a person forms a model of incoming discourse. Typically, one represents only those situations that are compatible with the truth of the incoming discourse and that are explicitly mentioned, possibly because of working memory limitations and because explicit true possibilities seem most relevant (Evans & Over, 1996; Sperber & Wilson, 1995). There is also evidence that people often exhibit minimalist tendencies, constructing a single model compatible with a proposition (Sloutsky & Goldvarg, 1999). Representations may thus deviate from logical prescriptions (e.g., a conditional with a false antecedent is true, but the conditional premise does not express these possibilities explicitly and they will probably not be represented). Second, the person draws an informative inference (not a repetition of what was given) by combining premise representations. Finally, the person

### Table 1

**Examples of Valid and Invalid Conditional Inference Forms**

<table>
<thead>
<tr>
<th>Validity</th>
<th>Argument form</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modus ponens</strong></td>
<td><strong>Modus tollens</strong></td>
</tr>
<tr>
<td>If then : If the weather is nice, Ed takes a walk.</td>
<td>If then : If the weather is nice, Ed takes a walk.</td>
</tr>
<tr>
<td>: The weather is nice.</td>
<td>: Ed does not take a walk.</td>
</tr>
<tr>
<td>: Ed takes a walk.</td>
<td>: The weather is not nice.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Valid</strong></th>
<th><strong>Invalid</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirming the consequent</td>
<td>Denying the antecedent</td>
</tr>
<tr>
<td>If then : If the weather is nice, Ed takes a walk.</td>
<td>If then : If the weather is nice, Ed takes a walk.</td>
</tr>
<tr>
<td>: Ed takes a walk.</td>
<td>: Ed does not take a walk.</td>
</tr>
<tr>
<td>: The weather is nice.</td>
<td>: The weather is not nice.</td>
</tr>
</tbody>
</table>
Table 2
Sample Modus Ponens (MP) and Affirming the Consequent (AC) Stor tales From Experiments 1 and 2

<table>
<thead>
<tr>
<th>Order of presentation</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP argument form</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Frank woke up on his couch after taking a long nap and realized that he didn’t know what time it was.</td>
</tr>
<tr>
<td>2</td>
<td>He thought that if it was cold outside, then it was night.</td>
</tr>
<tr>
<td>3</td>
<td>Still feeling sleepy, Frank arose to open a window.</td>
</tr>
<tr>
<td>Probe\textsuperscript{a}</td>
<td>He discovered that it was cold outside.</td>
</tr>
<tr>
<td>No inference\textsuperscript{a}</td>
<td>He wondered whether it was cold outside.</td>
</tr>
<tr>
<td>\textsuperscript{b} ** NIGHT **</td>
<td></td>
</tr>
</tbody>
</table>

AC argument form

| Inference\textsuperscript{a} | He discovered that it was cold outside. |
| No inference\textsuperscript{a} | He wondered whether it was cold outside. |
| \textsuperscript{b} ** NIGHT ** |        |

Sentence type

<table>
<thead>
<tr>
<th>Text sentences presented for recognition after the story\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical</td>
</tr>
<tr>
<td>Changed</td>
</tr>
<tr>
<td>Paraphrase</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Varied between participants. \textsuperscript{b} Presented in Experiment 1. \textsuperscript{c} Presented in Experiment 2.

searches for any counterexample serving as a model of the situation in which the premises are true but the inference is false. If no such model is found, the inference is accepted. Thus, the third step predicts that people are sensitive to the logical validity of conclusions, although logically incomplete representations tend to prevent ideal performance. The third step does not tie reasoning performance completely to comprehension processes because consideration of validity should be guided by this deliberate search process (see Evans, 2000).

Consider the story in the top third of Table 2 to examine the process of inference according to the mental models view. The mental models theory predicts that a conditional such as that in the second sentence will be represented in memory with tokens that correspond to the explicitly mentioned antecedent and consequent, respectively:

\[
\text{cold outside} \quad \ldots \quad \text{night}
\]

The ellipsis below these two tokens corresponds to other implicit possibilities compatible with the conditional’s truth but not made explicit.

How are MP inferences accomplished? When the participant learns that “it was cold outside,” as in the fourth sentence of the sample story, this model is also added to the participant’s representation of the input:

\[
\text{cold outside}
\]

This model matches up with the situation described in the model of the conditional premise because it repeats the conditional’s antecedent. In mental models terms, the effect of this match is to cancel the implicit possibilities and pick out the proposition that “it was night” (this is Step 2, the formation of a conclusion). This entire process accomplishes an MP inference without recourse to logical inference schemata.

AC inferences could plausibly occur by a similar process. Suppose that the participant instead encounters the story presented in the middle third of Table 2. The conditional premise will be represented in the same manner as in the previous example. When the participant processes the fourth sentence of this story, this model will be added to the discourse representation:

\[
\text{night}
\]

This model cancels the implicit possibilities and, by repeating the conditional’s consequent, matches the model of the conditional premise. It thus picks out the proposition that “it was cold outside” (again, Step 2 of the mental models theory). In this case, the entire process accomplishes an AC inference, and no additional assumptions need be invoked. Whether AC inferences occur during comprehension has almost completely been ignored by researchers, but we located one recognition study that suggested that these inferences do happen, at least as measured by false-alarm rates on these inferences (Franks, 1997).

This mental models account of model construction and inference is broadly consistent with a discourse processing perspective. The mental models theory suggests that premise representations are logically incomplete because of working memory limits and relevance considerations (Johnson-Laird & Byrne, 1991), but these factors also allow participants to consider as few possibilities as possible. In discourse comprehension, a major goal is to build a coherent representation (Gernsbacher, 1997; Kintsch, 1998; McKoon & Ratcliff, 1992; Zwaan & Radvansky, 1998), and minimizing possibilities to be considered could be one way to enhance coherence. Therefore, both MP and AC inferences were expected during discourse comprehension.

This prediction is also consistent with some specific findings in the discourse processing literature. MP inferences could be construed as predictive inferences, whereas AC inferences may be a type of bridging inference necessary to establish local coherence among adjacent propositions of a text (Haviland & Clark, 1974; Singer, 1994). Although the issue has been debated, the results of many studies have suggested that a person will draw a predictive inference when it is highly constrained, uniquely predicted by the text, and predictable on the basis of world knowledge (Fincher-Kiefer, 1995, 1996; Keefe & McDaniel, 1993; Klin, Guzmán, & Levine, 1999; McKoon & Ratcliff, 1986, 1989). MP inferences such as those based on the example story seem to meet these criteria, in that the conclusions are sensible in context and easily available because they are explicitly mentioned. Substantial evidence also exists that readers make bridging inferences (e.g., Fincher-Kiefer, 1995, 1996; Potts, Keenan, & Golding, 1988). AC inferences offer a way to connect the assertion of a conditional
with the assertion of the conditional’s consequent; such a connection is tantamount to drawing a bridging inference.

Finally, sentential connectives can serve as signals to make inferences that integrate separate propositions into coherent structures (Millis, Golding, & Barker, 1995; Millis & Just, 1994; Murray, 1997), and an if–then conditional may be one such signal. Logically, one should not draw AC inferences, but the cost of not doing so is to remain in a state of indeterminacy about the conclusion. If readers strive to maintain coherence, they may try to avoid this indeterminacy by committing the inference.

Unlike syntactic theories positing that reasoning is a part of comprehension, the mental models theory posits that reasoning, in addition to the construction of mental models, may involve a deliberate search for counterexamples. Many studies of deliberate reasoning find that the perceived necessity and sufficiency of a conditional’s antecedent for its consequent affect the inferences that reasoners will draw by making counterexamples more or less available in memory (Byrne, Espino, & Santamaria, 1999; Cummins, 1995; Markovits, 1993; Quinn & Markovits, 1998). However, searching for counterexamples could be too computationally demanding to occur automatically during comprehension (Kintsch, 1998), and these effects would not be expected. Mental models theory thus predicts that (a) searching for counterexamples is separate from comprehension, which involves only the first two steps of the reasoning process; (b) MP and AC inferences should therefore occur during comprehension; and furthermore (c) these inferences do not necessarily depend on a biconditional interpretation of conditionals.

Overview of Experiments

We used multiple methods because no single method of studying inference can give completely unambiguous results (e.g., Singer, 1994; Trabasso & Suh, 1993). Experiment 1 used an online priming task to assess the occurrence of MP and AC inferences during reading. Participants read short stories, presented under either an inference or no-inference condition, that supplied the premises for either MP or AC inferences, and they had to decide whether a single word presented immediately after each story had appeared in the story. For the critical stories, these words represented focal concepts of the inferences. If participants made the inferences while they read, they would be faster in responding to such words in the inference condition than in the no-inference condition, in which stories deviated from the MP and AC inference forms by not asserting a minor premise. Syntactic theories would predict that only MP inferences should be primed, whereas mental models theory would suggest that both MP and AC inferences should be primed.

In Experiment 2, participants read some of these stories, and after each they decided whether the information presented in each of three test sentences had been stated in the story. The critical test sentences represented conclusions to either MP or AC arguments that were not presented in the stories but were expected to be inferred by participants. Examples of arguments and critical test sentences are presented in Table 2. Note that conclusions to MP and AC arguments repeat one of the premises (If A then B. A. Therefore B). Therefore, if participants were asked merely whether B was stated, participants in both inference and no-inference conditions could have relied on memory of the conditional premise and overwhelmingly answered “yes.” For that reason, participants were asked whether the information in the arguments’ conclusions had been presented. If participants committed inferences as they read, they should have judged this information to have actually been presented in the inference condition. At the same time, in the no-inference condition the same sentences following stories that deviated from the MP and AC inference forms should have exhibited false alarms on these foils less frequently. Again, syntactic theories would suggest false alarms only for MP inferences, but mental models theory would predict false alarms for MP and AC inferences.

Finally, Experiment 3 was conducted to test the biconditional interpretation hypothesis that participants represent conditionals of the form If P then Q by maintaining in memory both a representation of that premise and a representation of the converse, If Q then P. Participants were thus tested for memory of both the original premise and its converse. We believed that an inability to distinguish between the two could corroborate the syntactic interpretation of AC inferences. We also felt that such a finding would weaken the mental models account.

Experiment 1

Method

Participants. A total of 56 introductory psychology undergraduates (33 women and 23 men) took part voluntarily in partial fulfillment of a course requirement. All participants were fluent English speakers.

Materials. There were 40 critical stories, each four sentences long. Each story contained in its second sentence a conditional premise formed from a noun–attribute pair. These pairs were selected from a preliminary study in which participants listed the first attributes that came to their mind in response to the noun. For each noun, the attribute selected was given by only 9.4% to 15.1% of participants, with a mean of 10.9%. These weakly associated propositions were used because in reasoning participants tend to avoid AC inferences for weakly associated premise information (e.g., “If it was cold outside, then it was night”), although committing these inferences for strongly associated premise information (e.g., “If it was dark outside, then it was night”), Janveau-Brennan & Markovits, 1999; Markovits, 1993; Markovits, Fleury, Quinn, & Venet, 1998; Quinn & Markovits, 1998).1

Using weakly associated premise information thus offers a more conservative test of the mental models hypothesis advanced here.

Each story was constructed so that the noun would denote the focal concept in either an MP or AC inference. To eliminate effects of a particular story on participants’ responses, we wrote each story in four versions: (a) MP-inference, (b) MP–no inference, (c) AC-inference, and (d) AC–no inference. An example of each version, using the noun–attribute pair night–cold, is displayed in Table 2. The MP–inference version contained second and fourth sentences that combined to fit the form: If attribute then noun; attribute, whereas the AC–inference version contained second and fourth sentences that combined to fit the form If noun then attribute; attribute. The MP–no inference and AC–no inference versions were identical to their inference counterparts except that the fourth sentence mentioned the attribute without asserting it. Hence, each story appeared in each of the four inference conditions: MP-inference, MP–no inference, AC–inference, and AC–no inference.

1 We also conducted an experiment in which items were strongly associated. The experiment used the identical procedure as Experiment 1 and yielded essentially the same results as the current experiment, in that during reading both MP and AC conclusions were primed in the inference condition but not in the no-inference condition.
The probe word was the same in all four versions of each story. Also, the fourth sentences for the inference and no-inference versions of each story were written to be as similar as possible and to contain many of the same concepts, to control for interlexical priming effects (Keenan & Jennings, 1995) that could inflate estimates of the priming effect because of inference form.

The 40 critical stories were randomly split into four groups of 10 stories each. Participants were also assigned randomly to one of four groups. Story groups and participant groups were combined in a Latin square design. Members of each participant group responded to one version of each critical story. Each participant thus responded to 10 stories in each of the four experimental conditions. Along with the 40 critical stories (all of which were expected to yield “yes” answers to the word probes), 40 filler stories were presented that were similar to the critical stories, in that they fit one of the four versions and were based on other noun–attribute pairs. For these fillers, the probe words did not appear in the stories, so they required “no” answers. Finally, there were 32 additional fillers, each four sentences long but lacking conditional premises of any kind. Half of these fillers required “yes” answers, and half required “no” answers. Participants thus received 112 stories in all. Participants also received 12 practice stories, which were not labeled as practice, before the 112 experimental stories; practice stories were not counted.

**Design and procedure.** The experimental design was a 2 (argument form: MP, AC) × 2 (premise version: inference, no-inference design). Both factors varied within subjects. Stimulus presentation and data collection were controlled by a PC, which ran Superlab Pro (Version 2.0; 1999) for Windows.

Participants were tested individually. Participants were told that they would read a series of stories on the computer and that after each story a word would be presented. The participant was to decide whether the word appeared in the story and to respond as quickly as possible by pressing one of two buttons (one for yes, another for no) on a four-button response box. Also, each participant was asked a yes–no comprehension question after each story (questions were included to encourage attentive reading; the answers did not pertain to any conditional inference that the participant could draw). Reasoning was never mentioned.

The participant read each story in a self-paced fashion; after reading one sentence, the participant used a response box button marked NEXT to clear the screen and display the next sentence. After the fourth sentence of the story, the participant’s press of the NEXT button cleared the screen and displayed the word probe bracketed by asterisks (e.g., **NIGHT**). After the participant responded, the screen cleared and a comprehension question appeared. As soon as the participant answered the question, the entire process repeated for the next story. All stimuli were presented in the same centered position on-screen in 14-point type. The overall duration of the experiment was approximately 45 min. The order of story presentation was randomized for each participant.

**Results and Discussion.**

In this and all subsequent experiments, an alpha level of .05 was adopted for statistical significance tests. In reported analyses, $F_1$ analyses collapsed across items and used error terms based on subject variability, and $F_2$ analyses collapsed across participants and used error terms based on item variability.

For analyses of reaction times, incorrect answers were discarded. For each participant, reaction times that were 2.5 or more standard deviations above the participant’s grand mean were also discarded, a procedure that removed 2.7% of all responses. Analyses of participants’ answers to the comprehension questions after each story indicated that participants were reading the stories attentively ($M$ errors = 11.6%, $SE = 0.6%$).

Participants’ mean reaction times in each condition, along with error rates, appear in Table 3. Mean reaction times were analyzed in 2 × 2 repeated measures analyses of variance (ANOVA), with argument form and premise version as within-subject factors. Results indicated a significant effect of premise version by subjects, $F_{1}(1, 55) = 29.99, p < .05$, and by items, $F_{2}(1, 39) = 13.71, p < .05$, with inference versions faster than no-inference versions.

The main effect of argument form and the Argument Form × Premise Version interaction were both nonsignificant, both $F_{1}(1, 55) < 1$, and both $F_{2}(1, 39) < 1$.

Finally, to examine the presence of any speed–accuracy tradeoffs that would compromise interpretation of the data, we conducted a 2 × 2 repeated measures ANOVA with the same factors as before on the numbers of errors that participants committed in each condition. Neither the main effects nor the interaction were significant in this analysis (all $Fs < 1$).

The results of this experiment have several implications. First, the commission of AC inferences along with MP inferences suggests that the participants were not applying logically valid inference schemata to input. Rather, the results of this experiment are consistent with the semantic position that people draw MP and AC inferences during comprehension during construction of mental models (Johnson-Laird & Byrne, 1991), with no search for counterexamples to conclusions. This is an important indicator of differences between comprehension and reasoning; it seems possible that conditional inferences during comprehension reflect only initial premise representation and combination. Search for counterexamples and any other reasoning strategies that participants might possess do not seem to have been applied in this task.

We also wished to examine whether participants actually encode the MP and AC inferences into their memory representations of the stories. We predicted that if such encoding did not occur, participants would not agree that the stories contained these inferences if explicitly asked after reading the stories. On the other hand, we believed that if participants agreed that these inferences had been presented, we would have more confidence that the inferences were actually being made. To examine this possibility, we conducted Experiment 2.

**Experiment 2.**

**Method.**

Participants. The participants were 48 undergraduate introductory psychology students (19 women and 29 men) who took part voluntarily in partial fulfillment of a course requirement.
Materials. There were 20 critical stories selected at random from those used in Experiment 1; each story had four versions (MP-inference, MP-no inference, AC-inference, and AC-no inference). However, participants did not respond to probe words after each story. Instead, they first answered a yes–no comprehension question (included to encourage participants to read attentively). Then, they received three test sentences—one of each of the following types:

1. Critical test sentences represented either MP or AC conclusions (depending on condition).
2. Paraphrase test sentences presented the concepts of one of the sentences from the story with wording altered from the original.
3. Changed test sentences presented the concepts from one of the story sentences, but one proposition was changed by inserting a noun phrase that differed from the original.

The important test sentences were the critical sentences; the paraphrase and changed sentences were added because they should have elicited “yes” and “no” responses, respectively, and thus prevented the formation of a “yes” or “no” response bias by participants. Examples of stories and test sentences appear in Table 2.

The 20 critical stories were split randomly into four groups of 5. Each participant was also assigned randomly to one of four groups. Participant groups and story groups were combined in a Latin square design, so that each participant responded to one version of each story. Each participant received 5 stories in each of the four experimental conditions. Each participant also responded to 15 nonconditional filler stories that had been selected from the nonconditional fillers used in Experiment 1, making a total of 35 stories. These filler items were included to monitor honest reading of sentences by participants. Following each filler, participants also received three test sentences—a paraphrase, a changed sentence, and a critical sentence. The critical sentences represented nonlogical elaborative and bridging inferences that readers might have made, rather than inferences consistent with propositional logic. The three test sentences were presented in random order after each story (for both critical and filler stories).

Design and procedure. The experimental design included the within-subject factors of argument form (MP, AC) and premise version (inference, no inference). Stimulus presentation and data collection were controlled by a PC that ran Superlab Pro 2.0 for Windows.

Participants were tested individually. Participants were told that they would read a series of stories on the computer and that after each story they would answer a question and then decide, for each of three test sentences, whether the information in each test sentence had been presented in the story. An example story was presented at this point, and the participant was given two test sentences about this story (one paraphrase and one changed) along with feedback about each. The feedback demonstrated that the participant should make decisions on the basis of how well the information in each test sentence matched the content of the original story, rather than how well the test sentences matched the original stories’ exact wording. Reasoning was never mentioned.

Participants read stories in a sentence-by-sentence process identical to that of Experiment 1. This process repeated until the fourth sentence; after the participants responded, the comprehension question appeared. The participants responded to the question by pressing the Z key for yes and the / key for no. After this, each test sentence was presented one at a time in random order. The participants pressed Z for yes if they believed that the information in the sentence had been presented in the story, and // for no if they believed that the information in the sentence had not been presented. This process then repeated for each story. Order of story presentation was randomized for each participant.

Results and Discussion

Inspection of the answers to the comprehension questions indicated that participants appeared to read the stories attentively (M errors = 11.2%, SE = 1.9%). Participants also tended correctly to reject the changed test sentences (M errors = 6.2%, SE = 1.0%) and to accept the paraphrase test sentences (M errors = 10%, SE = 1.0%).

Mean acceptance rates for the critical test sentences appear in Table 4. Participants’ numbers of acceptances of critical test sentences in each experimental condition were analyzed in 2 × 2 repeated measures ANOVAs with argument form and premise version as within-subject factors. These analyses revealed a premise version effect that was significant by subjects, $F_{1}(1, 47) = 67.89, p < .05$, and by items, $F_{2}(1, 19) = 199.87, p < .05$, such that inference premise versions ($M = 60\%, SE = 4.6\%$) led to more acceptances than no-inference premise versions ($M = 22.3\%, SE = 3.6\%$). The main effect of argument form and the Argument Form × Premise Version interaction were not significant in either analysis (all $Fs < 1$).

Results are largely analogous to the pattern found in Experiment 1. Whenever a critical test sentence followed either an MP or AC argument of which the critical sentence was the conclusion, participants tended to agree that they had read the information in the critical sentence. Whenever a critical sentence was presented after a story that did not fit either form, participants tended to say that they had not read the information in it. These results further support the semantic explanation of conditional processing that predicts the frequent occurrence of AC inferences.

One further experiment was necessary to examine another putative explanation of participants’ tendency to make AC inferences found in Experiments 1 and 2. It was possible that participants, on processing conditionals of the form If $P$ then $Q$, simultaneously formed representations of If $P$ then $Q$ and its converse If $Q$ then $P$. This amounted to the biconditional interpretation alleged in many studies, an interpretation that rendered AC inferences valid and upheld the syntactic position (see Braine & O’Brien, 1991; Romain et al., 1983). If this interpretation was correct, we hypothesized, participants tested for their memory of the conditional premise immediately after reading a story might have difficulty discriminating between the original, If $P$ then $Q$, and the converse, If $Q$ then $P$, because they would tend to false alarm on the converse. If they tended not to commit such false alarms, then this biconditional explanation would not be supported.

Experiment 3

Experiment 3 tested the biconditional interpretation outlined above. In this experiment, we used strongly associated premise information as well as weakly associated information. Although we ran versions of Experiments 1 and 2 that used strongly associated premise information and essentially replicated those experiments’ results (see Footnotes 1 and 2), we felt that strongly associated items should be included to give a fair test to the biconditional hypothesis. These items could be more likely than weakly associated conditional premises to elicit biconditional in-
interpretation, therefore inclusion of the strongly associated items could represent a more sensitive test of the hypothesis.

**Method**

**Participants.** Twenty-two introductory psychology undergraduates (15 women and 7 men) took part in the experiment voluntarily in partial fulfillment of a course requirement. All were fluent English speakers.

**Materials.** The MP-inference version of each of the stories in Experiment 2 was used. In addition, a second version of each story containing a conditional premise formed from a strongly associated noun-attribute pair was also constructed; these strongly associated nouns and attributes were identified in the same preliminary study described in Experiment 1. For example, although night and cold are associated weakly, night and dark are associated strongly. There were thus 40 critical stories in all (20 with strongly associated antecedents and consequents and 20 with weakly associated antecedents and consequents). The 15 filler stories from Experiment 2 were also used.

For each story, three memory test sentences were constructed for evaluation after the story. Included were one of each of the following types:

1. **Original** sentences from each critical story contained the conditional if P then Q premise (e.g., “If it was cold outside, then it was night.”).
2. **Foil** sentences were identical to the originals, except conditional premises were rearranged into their converses, If Q then P (e.g., “If it was night, then it was cold outside.”)
3. **Control** sentences were identical to the originals, except one noun or noun phrase in each conditional premise was changed into another noun phrase that was also sensible for that sentence (e.g., “If it was cold outside, then it was early morning.”). Each such change was only a one- or two-word alteration from the original.

Three memory test sentences were constructed for each filler story, on the basis of sentences that included other nonlogical connectives, such as because, before, and after. These filler items were introduced to monitor the honest reading of sentences. After 5 of the fillers, original, foil, and control sentences analogous to those for the critical stories were presented. After another 5 fillers, the original sentence was presented twice along with the foil. After the other 5 fillers, the foil was presented twice along with the control. Fillers were manipulated in this way so that participants could not learn to expect one old sentence after each story.

**Design and procedure.** The design included the factors of memory sentence (original, foil, and control) and content type (strongly associated vs. weakly associated). Memory sentence type was varied within subjects and content type was varied between subjects, with 11 participants assigned to each content type group (each participant thus received 35 stories: 20 critical and 15 filler).

Participants were instructed to read the stories carefully because they would be tested for memory after reading and answering a question about each story. Participants were told for the memory test to answer “yes” only if they thought that a test sentence was exactly the same as a sentence from the story. They were also told that there could be zero, one, or more than one “yes” sentence after each story. Stimulus presentation and data collection were controlled by a PC, which ran Superlab Pro 2.0 for Windows. The method of story presentation was the same sentence-by-sentence, self-paced method used in previous experiments. Following each story, participants answered a yes–no comprehension question using the Z key to answer yes and the // key for no. These questions were identical to those used in the above reported experiments. After the question, the three memory items for that story were presented one at a time in a random order; participants responded to each using the same two keys. Then the process repeated for the next story. Order of story presentation was randomized for each participant.

**Results and Discussion**

Participants seemed to read the stories attentively, as shown by high accuracy in answers to the comprehension questions (M errors = 9.9%, SE = 1.7%). Inspection of participants’ responses to memory sentences indicated that control sentences were virtually never accepted; the mean acceptance rates in the strongly associated and weakly associated conditions were 1.8% and 0.5%, respectively. The control condition was thus dropped. The numbers of acceptances made by each participant to memory test sentences of the remaining types (original and foil) were analyzed in a 2 (content type) × 2 (memory sentence) mixed-participants ANOVA, with repeated measures on the latter factor. Acceptance rates for the four conditions are presented in Table 5. The analysis revealed a main effect of memory sentence, F(1, 20) = 385.47, p < .05, with originals (M = 87%, SE = 2.1%) accepted significantly more often than foils (M = 12.3%, SE = 2.2%). The main effect of content type was not significant, F(1, 20) < 1. The Content Type × Memory Sentence interaction was also significant, F(1, 20) = 4.36, p < .05. Follow-up simple effect analyses indicated that no difference existed in acceptance rates for original sentences as a function of content type, F(1, 20) = 1.71, p > .2. However, acceptance rates for foil sentences did differ as a function of content type, F(1, 20) = 5.15, p < .05, with foils accepted more often in the strongly associated condition than in the weakly associated condition. Given that strongly associated items are by definition linked in semantic memory, it is not surprising that more confusion should occur in that condition. Both the If P then Q original and the If Q then P converse may seem plausible, whereas the converse for many weakly associated items is not likely to be plausible.

However, acceptance rates for foils were well below chance, t(10) < −10, ps < .05, in both the strongly associated and weakly associated conditions. Indeed, no participant accepted more than 8 of 20 foils in the strongly associated condition or more than 7 of 20 foils in the weakly associated condition, with average acceptance

<table>
<thead>
<tr>
<th>Content</th>
<th>Test sentence</th>
<th>Original: If P then Q</th>
<th>Foil: If Q then P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly associated</td>
<td>84.1 (3.6)</td>
<td>17.3 (3.3)</td>
<td></td>
</tr>
<tr>
<td>Weakly associated</td>
<td>90 (2.7)</td>
<td>7.3 (3.0)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Standard errors are in parentheses.
below 18% for strongly associated items and below 8% for weakly associated items. Acceptance rates for originals were well above chance in both conditions, both ts(10) > 9, ps < .05. The biconditional interpretation, as measured in this experiment, seems to be too infrequent and unreliable to account for systematic priming observed in Experiment 1 and for the large proportions of false alarms observed in Experiment 2. Even when items were strongly associated, biconditional errors were quite infrequent. This interpretation is thus not a likely explanation of the occurrence of AC inferences for the previous experiments.

General Discussion

The reported experiments examined the types of conditional inferences that people tend to make during comprehension, to help identify the representations of conditional premises that participants form prior to deliberate reasoning processes. Data from Experiment 1 indicate that the MP and AC premise forms prime their conclusions during reading even when the conditional’s antecedent and consequent are weakly associated in memory. Results of Experiment 2 suggest that these MP and AC inferences are encoded into memory, again even when the conditional’s antecedent and consequent are weakly associated in memory. Recall that, as mentioned in Footnotes 1 and 2, these effects were replicated with strongly associated items.

Finally, results of Experiment 3 suggest that participants are not drawing both MP and AC inferences because they represent both a conditional and its converse in memory; participants infrequently falsely recognized conditionals’ converses but reliably accepted the conditionals themselves. The results with MP replicate previous studies (Franks, 1996, 1997; Lea, 1995; Lea et al., 1990), whereas the AC results are more novel (Franks, 1997, being the only previous report). In fact, the demonstration that AC inferences seem to occur online is a completely novel phenomenon.

These findings are compatible with predictions derived from the semantic approach, particularly mental models theory of reasoning, but they run counter to predictions of syntactic theories of reasoning. Syntactic theories predict that (a) reasoning is largely a component of comprehension; (b) MP inferences should occur during comprehension; and (c) whenever AC inferences do occur, they should occur because people tend to represent both a conditional and its converse. None of these predictions was supported by the reported experiments. Predictions of the mental models theory are that (a) the last step in deliberate reasoning, searching for counterexamples, is separate from comprehension and (b) MP and AC inferences should occur during comprehension. These predictions were supported in these experiments. Therefore, the reported findings raise important implications for (a) syntactic and semantic theories of conditional processing and (b) the relationship between comprehension and reasoning.

Syntactic theories propose that people automatically extract the propositional form of conditionals and apply the MP inference schema to these forms, whenever possible (Braine, 1990; Braine & O’Brien, 1991; Lea, 1995; Lea et al., 1990). Given that the MP schema and all inference schemas postulated by syntactic theories are valid, the existence of online AC inferences suggests that these theories are incomplete. Furthermore, the existence of online AC inferences suggests that people may not automatically parse discourse into logically valid syntax; such parsing may occur only whenever people are asked to reason (and even then the parsing may be nonveridical). Indeed, some researchers argue that logical reasoning performance depends on the deliberate application of one or more strategies rather than on automatic comprehension processes (Evans, 2000; Johnson-Laird, Savary, & Bucciarelli, 2000; see also Stanovich, 1999).

Could the biconditional interpretation hypothesis or the invited inference hypothesis account for the frequent commission of AC inferences during reasoning (e.g., Evans et al., 1993) and comprehension? Recall that the biconditional interpretation hypothesis suggests that whenever presented with an If P then Q statement, participants maintain its converse If Q then P in working memory. The invited inference hypothesis suggests that whenever presented with an If P then Q statement participants believe that If not-P then not-Q is also the case (Geis & Zwicky, 1971). Neither proposal seems capable of accounting for the observed invalid inferences. According to the biconditional interpretation hypothesis, someone hearing If P then Q is led to believe that if Q then P is also the case. In this case Q, although a consequent in If P then Q, is an antecedent in If Q then P. Therefore, according to this interpretation people make an MP inference in If Q then P, which appears as if it were an AC inference in If P then Q. Our results do not support such an explanation. Although the proportions of false alarms evidencing invalid AC inferences in Experiment 2 were approximately 60%, the proportion of false alarms on If Q then P converses in Experiment 3 was below 8% for weakly associated items, and this proportion did not surpass 18%, even for strongly associated items. Therefore, it seems that the biconditional interpretation cannot account for frequent AC inferences during comprehension.

What about the invited inference hypothesis, that someone hearing If P then Q is led to believe that if not-P then not-Q is also the case? One possibility is that when presented with Q, which is the negation of the consequent in If not-P then not-Q, people may conclude that P is also the case, which is the negation of the antecedent in If not-P then not-Q. Such an inference would conform, under the invited inference hypothesis, to a valid MT schema (see Table 1). However, the syntactic approach posits no MT schema, and reasoners typically draw MT inferences much less frequently than MP inferences (e.g., Evans et al., 1993; Rips, 1995). Invited inferences thus will not account for the observed AC inferences. In sum, it appears unsatisfactory to explain away invalid inferences in reasoning and comprehension as being due to invited inferences or representing conditionals as biconditionals.

Semantic theories, in particular the mental models theory (Johnson-Laird & Byrne, 1991), predict that the representation of a conditional formed during comprehension consists mainly of a conjunction of the antecedent and consequent. If further input supplies the premise needed for either an MP or AC inference, this inference is likely to be drawn during comprehension because both inferences can be based on the representations formed during comprehension. The reported findings also indicate that the initial construction of mental models during comprehension of conditional premises is distinct from searching for counterexamples, which is likely to be a later phase in the course of deliberate reasoning. In reasoning, participants tend to avoid AC inferences for weakly associated premise information, although committing these inferences for strongly associated premise information.
In our experiments, however, participants committed AC inferences with weakly associated items. Directly presenting counterexamples may lead to their use in deliberate conditional reasoning tasks (Byrne, 1989; Byrne et al., 1999; Rumain et al., 1983), as could the use of conditional premises whose semantic and pragmatic content strongly suggests counterexamples to inferences (e.g., Cummins, 1995; Markovits, 1993; Markovits et al., 1998). However, there is currently no evidence that searching for counterexamples or any other deliberate reasoning strategy comes into play when participants are not explicitly requested to reason.

Turning to the relationship between comprehension and reasoning, although some researchers argue that much of propositional reasoning is essentially part of language comprehension and thus procedural in nature (Braine, 1990; Braine & O’Brien, 1991; Lea, 1995; Macnamara, 1986), the findings of the current study do not support this argument. Deductive reasoning is not exhausted by comprehension, and the former but not the latter is a deliberate process requiring strategic thinking to reach solutions (see Evans, 2000; Stanovich, 1999, for similar arguments). This strategic thinking may take the form of searching for counterexamples to conclusions or even inferences based on suppositions; different individuals may use different strategies (Johnson-Laird et al., 2000; Stanovich, 1999). Comprehension, on the other hand, is a rapid automatic process that seeks to establish a coherent representation of a text (Gernsbacher, 1997; Kintsch, 1998; Zwaan & Radvansky, 1998), and one could argue that both MP and AC inferences accomplish this goal. These inferences maintain coherence by allowing a person to minimize the number of possibilities to be considered, a factor to which people seem to be sensitive (Johnson-Laird & Byrne, 1991). Thus, establishing coherence may yield results that are different from those that would obtain if a person were reasoning deductively. These phenomena are quite compatible with the process of inference proposed by mental models theory. On the basis of the results of the reported experiments, the hypotheses of mental models theory about initial representations appears to be more consistent with the types of inferences drawn during comprehension.

The suggestion is not that people cannot reason; instead, the point is that more of the reasoning process may be deliberate than syntactic theorists seem to believe. Inferences made during comprehension may be only precursors to those made during deliberate reasoning. Thus, strategies that have been proposed to account for conditional reasoning in straightforward deductive tasks, such as a Bayesian information-maximization strategy (Oaksford, Chater, d’Ydewalle (Eds.), Deductive reasoning and strategies (pp. 1–22). Mahwah, NJ: Erlbaum.


Received November 8, 2000
Revision received May 10, 2001
Accepted May 16, 2001