Reasoning From Double Conditionals: The Effects of Logical Structure and Believability

Carlos Santamaría

Universidad de La Laguna, Tenerife, Spain

Juan A. García-Madruga

UNED, Madrid, Spain

Philip N. Johnson-Laird

Princeton University, New Jersey, USA

We report three experimental studies of reasoning with double conditionals, i.e. problems based on premises of the form:
    If A then B.
    If B then C.
where A, B, and C, describe everyday events. We manipulated both the logical structure of the problems, using all four possible arrangements (or "figures" of their constituents, A, B, and C, and the believability of the two salient conditional conclusions that might follow from them, i.e. If *A then C*, or If *C then A*. The experiments showed that with figures for which there was a valid conclusion, the participants more often, and more rapidly, drew the valid conclusion when it was believable than when it was unbelievable. With figures for which there were no valid conclusions, the participants tended to draw whichever of the two conclusions was believable. These results were predicted by the theory that reasoning depends on constructing mental models of the premises.

INTRODUCTION

Human performance in deductive reasoning appears to depend on two main variables: logical structure and semantic content (see e.g. Evans, Newstead, & Byrne, 1993). Most theories treat these factors as independent. Thus, on the one

hand, many studies have demonstrated the effects of structure on deductive reasoning (see e.g. Rips, 1994), and on the other hand, many studies have demonstrated the effects of content on deductive reasoning (see e.g. Wilkins, 1928). Content likewise engages reasoners' beliefs, and they too influence performance—effects that have been usually treated as a bias on the interpretation of the premises or as a response bias (see Henle & Michael, 1956; Revlin & Leirer, 1978). If reasoning is a formal process, then in principle it is impossible for content or beliefs to affect the *process* of reasoning, which, by definition, is a purely formal syntactic business. In contrast, if reasoning is a process of envisioning models of situations—as proponents of the mental model theory suppose (see e.g. Johnson-Laird & byrne, 1991)—then content may have direct effects on the process itself. Reasoners may, for example, search more assiduously for alternative models of the premises if their initial model yields an unbelievable conclusion from believable premises (see Oakhill & Johnson-Laird, 1985; Oakhill, Johnson-Laird, & Garnham, 1989). Our aim in this paper is to examine these two factors: the logic of the problem, and the believability of the possible conclusions.

Nearly all studies of the effects of believability on reasoning have concerned syllogisms, such as:

> All of the Frenchmen are wine drinkers.
> Some of the wine drinkers are gourmets.
> ∴   Some of the Frenchmen are gourmets.

Many of the early studies were methodologically flawed, but recent research has established three main phenomena concerning logic and believability in the case of syllogisms (see e.g. Evans, Barston, & Pollard, 1983). First, reasoners accept more valid conclusions than invalid conclusions. Second, they accept more believable than unbelievable conclusions. Many people, for example, draw the conclusion given the syllogism above (see Oakhill et al., 1989). It is invalid, but highly believable. Third, there is an interaction between the two factors: the effect of believability is greater on invalid inferences than on valid ones. In particular, the difference in the percentages of believable and unbelievable conclusions that reasoners accept is greater for invalid than for valid syllogisms. This interaction occurs when the invalid conclusions are logically consistent with the premises, but do not follow necessarily from them. When invalid conclusions are inconsistent with the premises, then the interaction disappears (Newstead, Pollard, Evans, & Allen, 1992). The effect of beliefs is less marked when participants have to draw their own conclusions (Barston, 1986), but the same interaction between logic and believability does occur (see e.g. Oakhill et al., 1989).

Theorists have proposed at least three alternative accounts for the phenomena. According to the *misinterpreted necessity* hypothesis (Evans, 1989; Evans, 1983), which was inspired by Dickstein (1980, 1981), reasoners fail to grasp that a conclusion that is merely consistent with the premises does not thereby follow necessarily from them. Hence, reasoners try to show that a conclusion is necessarily true, or necessarily false, given the premises. If it is necessarily true, then it follows from the premises; and if it is necessarily false, then it does not follow from the premises. But, if reasoners fail to reach either of these decisions, then they fall back on believability, and tend to accept believable conclusions and to reject unbelievable conclusions.

According to the *selective scrutiny* hypothesis (Evans, 1989; Evans et al., 1983), reasoners start by assessing the believability of the given conclusion. If it is believable, they tend to accept it; if it is not believable, they assess whether or not it follows from the premises, and make their decision accordingly. As Evans (1989, p.78) points out, selective scrutiny is consistent with other sorts of bias. It is also consistent with two other results: the failure of instructions that clarify logical necessity to reduce the interaction between logic and believability (Barston, 1986), and the failure of logical difficulty to enhance the interaction (see the unpublished study carried out by Evans and Pollard, which is described by Evans, 1989, pp.77–78).

According to the *mental model* theory, the phenomena arise from the effects of believability on the search for alternative models. Reasoners tend to "satisfice", that is, if their initial models of the premises support a believable conclusion, they will look no further; likewise, they will look no further if their initial models are inconsistent with a given conclusion; but if their initial models support an unbelievable conclusion, then they will search assiduously for an alternative model of the premises that refutes the unbelievable conclusions. Because individuals often reject unbelievable valid conclusions from syllogisms that require only one model, the model theory adds a filter on conclusions. Evans and Over (1996, p.112) argue that this hypothesis is both *ad hoc* and unparsimonious. Unfortunately, the result it is designed to explain is equally embarrassing for both selective scrutiny and misinterpreted necessity (see Evans, 1989, p.75–76). The conclusion is unbelievable, and so, according to selective scrutiny, reasoners should base their decision on whether or not it follows from the premises. It does follow from the premises, and so reasoners should accept it—but they don't. Likewise, the conclusion follows from the premises, and so, according to misinterpreted necessity, it should be accepted. All three accounts, therefore, require some additional principle to explain the rejection of valid unbelievable conclusions. Hence, we believe that the model theory provides a possible underlying mechanism for selective scrutiny (see also Barston, 1986), but, as we shall see, the two approaches diverge in certain cases.

DOUBLE CONDITIONALS

In order to re-examine logic and believability, we decided to examine a different sort of inference instead of syllogisms. Our test-bed was a set of inferences based on "double" conditionals (or "sorites" as they are sometimes known, see Evans et al., 1993), i.e. pairs of conditional premises, such as:

> If Concha studies, then she passes the exam.
> If Concha passes the exam, then she finishes the course.

The participants' task was to draw a conclusion by filling in the blanks in a sentence skeleton in the form of a conditional. The previous premises, for example, support the valid conclusion:

> If Concha studies, then she finishes the course.

This conclusion is valid because it must be true given that the premises are true. The inference has the following structure, or figural arrangement of the propositions, which we will designate *figure 1* (following the conventions for designating figures of syllogisms used by Johnson-Laird and his colleagues, e.g. Johnson-Laird & Bara, 1984):

> Figure 1:   If A–B
> If B–C
> ∴   If A–C

where A and C designate the two "end" propositions, and B designates the "middle" proposition common to both premises. In this figure, the conclusion:

> If C–A

is invalid. Here, and throughout the paper, such assessments of validity and invalidity depend on the assumption that the premises are interpreted as "one-way" conditionals rather than as bi-conditionals equivalent to: *If, and only if, A then B*; *If, and only if, B then C*. If we switch round the order of the two premises:

> If Concha passes the exam, then she finishes the course.
> If Concha studies, then she passes the exam.
> ∴. If Concha studies, then she finishes the course.

the inference is in a new figure:

> Figure 2:   If B–A
> If C–B
> ∴   If C–A

In this figure, the conclusion:

> If A–C

is invalid. Premises in the figures:

    Figure 3:   If A–B
                 If C–B

and:

    Figure 4:   If B–A
                 If B–C

do not yield valid conclusions interrelating the end propositions. Hence, we will refer to these two figures as "indeterminate", and to the two figures that yield valid conclusions as "determinate".

There are well known figural effects in syllogistic reasoning (Dickstein, 1978; García-Madruga, 1982; Johnson-Laird & Steedman, 1978). In particular, premises of the form: A–B, B–C, tend to yield conclusions of the form: A–C; whereas premises of the form B–A, C–B, tend to yield conclusions of the form C–A, regardless of whether the conclusions are valid or invalid. The interpretation of these effects is controversial, but in our view they reflect the order in which the premises are processed in order to integrate them; that is, the conclusions are read off from models in the same order as the premises were read in to working memory (the "first in, first out" principle, see Broadbent, 1958). If so, similar figural effects should occur with double conditionals.

In double conditionals, the believability of the conclusions can be manipulated independently from their structure and from the believability of their premises. For example, a problem in figure 1, *If A–B If B–C*, such as:

    If Concha studies, then she passes the exam.
    If Concha passes the exam, then she finishes the course.

has a valid conclusion that is believable:

    If Concha studies, then she finishes the course.

and a converse conclusion that is invalid and unbelievable:

    If Concha finishes the course, then she studies.

The same figure, however, can have a different semantic content that yields the opposite effects. For example, in the case of:

If Marta is hungry, then she takes an afternoon snack.
If Marta takes an afternoon snack, then she has a light dinner.

the A–C conclusion:

If Marta is hungry, then she has a light dinner.

is valid, but unbelievable. In contrast, the C–A conclusion:

If Marta has a light dinner, then she is hungry.

is invalid, but believable. It is tempting to suppose that if a valid conclusion is unbelievable, then some of the premises, or the combination of the premises, must be difficult to believe. However, the supposition is false. Consider, again, the premises:

If Marta is hungry, then she takes an afternoon snack.
If Marta takes an afternoon snack, then she has a light dinner.

Both premises, taken together or in isolation, are believable. Yet, the valid conclusion that they yield is not believable:

If Marta is hungry, then she has a light dinner.

The effect arises because the middle term refers to an event that is the vital causal link in the chain from antecedent to consequent. Without the middle term, however, the conclusion violates the normal relation between hunger and dinner. Using this type of example, certain philosophers argue that the ordinary conditional is not transitive—a view that, of course, leads to a very different analysis from ours, and that does not appear to account for the inferences made by the participants in our experiments. In our analysis, transitivity is valid for the ordinary conditional, but without fully explicit models of the premises and the conclusion, people can sometimes find the conclusion unbelievable when the premises are believable. (See Johnson-Laird & Savary, 1996, on the illusory inferences of people who do not make their models of premises and conclusions fully explicit.)

If we switch round the order of the two premises, then the result is an analogous pair of problems in figure 2: *If B–A If C–B*, which have either a believable valid conclusion and an unbelievable invalid conclusion, or else an unbelievable valid conclusion and a believable invalid conclusion. For premises in the indeterminate figures, both *If A–C* and *If C–A* conclusions are invalid, but it is possible to manipulate the content of the problems so that one of these conclusions is believable and the other is unbelievable.

In sum, we can produce inferences in which each pair of premises is believable, and for each determinate figure there is either a valid believable and an invalid unbelievable conclusion or else a valid unbelievable and an invalid believable conclusion. Likewise, for each indeterminate figure there is a believable and an unbelievable conclusion.

The model theory proposes that conditional assertions of the form:

If A then B

are normally represented by two models. The first model makes explicit the case where the antecedent A is true, and so the consequent B is also true. Reasoners do not normally, however, represent explicitly the case or cases that may hold if the antecedent A is false. The fundamental representational assumption of the theory is indeed that individuals normally represent only what is true, not what is false. Hence, they construct only an implicit model to represent the cases where the antecedent is false. The two models of the conditional are accordingly:

a   b
   …

where "a   b" denotes a model of the state of affairs described by the propositions A and B, and the ellipsis denotes the wholly implicit model. Reasoners need to make a "mental footnote" that the explicit model exhausts the cases where A is true, i.e. A cannot occur in the cases represented by the wholly implicit model. As the explicit model does not exhaust the cases where B is true, B may occur in the cases represented by the wholly implicit model. Johnson-Laird and Byrne (1991) used a special notation of square brackets for these mental footnotes, but we shall forego this notation. If reasoners retain the mental footnotes, then they can flesh out the models of the conditional to make them fully explicit:

    a   b
 ¬ a   b
 ¬ a   ¬ b

where "¬" denotes negation. The bi-conditional has the same initial models as the conditional. The only difference is that the mental footnotes for a bi-conditional specify that the explicit model exhausts both the cases where A is true and the cases where B is true. If reasoners retain these mental footnotes, then they can flesh out the models to make them fully explicit:

    a   b
 ¬ a   ¬ b

Much recent evidence suggests that reasoners rapidly forget mental footnotes and are accordingly susceptible to a variety of illusory inferences (see e.g. Johnson-Laird & Savary, 1996).

The theory predicts that a double conditional in figure 1:

If A then B
If B then C

has two initial models:

a   b   c
      …

where the first model is of the case where A, B, and C, are true, and the ellipsis denotes the other possibilities, which reasoners do not normally make explicit. These models yield the conclusion:

If A then C.

If this conclusion is believable, then reasoners will go no further. If this conclusion is unbelievable, however, then they will think again. They may scan the models in the opposite direction, or construct models based on the converses of the premises, and as a result draw the invalid conclusion:

If C then A

especially if it is believable. Alternatively, they may reject the explicit model:

a   b   c

They will then be left only with the wholly implicit model from which nothing follows, and so they will respond that there is no valid conclusion (see Johnson-Laird & Byrne, 1991, for a similar explanation of this response in the case of *modus tollens*).

The theory predicts similar models for premises in figure 2:

If B then A
If C then B

except that participants should be more likely to build their models starting from the second premise:

c   b   a
      …

Once again, the theory makes the same predictions about the effects of believability. But, the biases should occur to a lesser degree than in figure 1, because they arise in figure 2 only as a result of re-ordering the premises.

The initial models of the indeterminate premises in figures 3 and 4 can be of either form:

    a   b   c
         …

or:

    c   b   a
         …

and so yield conclusions of either form. If a conclusion supported by the mental models is unbelievable, the participants can reject it and consider the converse models. In order to discover that either conclusion is invalid given a "one-way" interpretation of the conditionals, reasoners need to flesh out their models fully explicitly. Thus, the fully explicit models for premises in figure 3:

    If A then B
    If C then B

are :

       a   b   c
       a   a   ¬ c
    ¬ a   b   c
    ¬ a   b   ¬ c
    ¬ a   ¬ b   ¬ c

which do not support either of the two conclusions: *If A–C*, or *If C–A*.

There is a critical distinction between studies of syllogisms and our studies of double conditionals. With syllogisms, the content of the *premises* was manipulated so that some syllogisms support a believable initial conclusion whereas other syllogisms support an unbelievable initial conclusion. The results showed that the believability of the conclusions had more effect when they were invalid than when they were valid. With double conditionals, however, *one and the same set of believable premises* supports both a believable conclusion and a converse unbelievable conclusion. Our present studies were therefore designed to test the three predictions of the model theory that we have described.

1. For the determinate figures 1 or 2, reasoners should tend to draw the valid conclusion, especially when it is believable, but when it is unbelievable they

should tend either to draw the converse believable conclusion or to respond that there is no valid conclusion.

2. These biases should be greater for figure 1 (*If A–B If B–C*) than for figure 2 (*If B–A If C–B*), because they arise in the latter case as a result of re-ordering the premises.

3. For the indeterminate figures 3 or 4, reasoners should tend to draw whichever of the two conclusions is believable.

EXPERIMENT 1

The aim of the experiment was to test the foregoing three predictions. The participants' task was to draw a conditional conclusion interrelating the end propositions of the two premises by filling in a skeleton sentence.

METHOD

Design.   The participants carried out eight different sorts of inference which were presented in a different random order to each participant. The eight sorts of inference derived from the four different figures:

1. If A–B   If B–C
2. If B–A   If C–B
3. If A–B   If C–B
4. If B–A   If B–C

and two different assignments of believable contents that made either the *If A–C* conclusion believable and the *If C–A* conclusion unbelievable, or vice versa. Each pair of premises was believable, and no participant received two problems with the same content. The particular 8 problems for an individual participant were selected by rotation from the set of 24 so that each of the 24 problems was used equally often in the experiment as a whole.

Participants.   32 psychology students at the University of La Laguna, Tenerife, carried out the experiment in order to receive credit for a course. They had not previously participated in an experiment on logical reasoning.

Materials.   We constructed the 24 problems, which were in Spanish, from believable pairs of conditionals (see the examples in the Introduction). All the problems shared similar pragmatic features: they described a course of actions carried out by an individual, with one state of affairs leading to another. Each problem yielded a believable conclusion (e.g. If John screams, then he gets help), and a converse unbelievable conclusion that violated the usual causal relation (If John gets help, then he screams).

In order to ensure that the conclusions varied in believability in the required way, we carried out a normative study with a separate group of 32 students from the same university class as the experimental participants. Their task was to select the more believable conclusion from each of 12 pairs of conditionals (of the form: *If A–C*, and *If C–A*, corresponding to the conclusions in the experiment proper). The order of the conditionals within the pairs was counterbalanced, so that half of the participants received a given conditional first in the pair, and the other half of the participants received it second in the pair. The 12 pairs were presented in a different random order to each participant: 96% of the participants' judgements were in accordance with ours, and there were no reliable difference over the different pairs (Friedman test: $\chi^2 = 1.637$, d.f. $= 11$, $P > .99$).

Procedure.   The problems were presented to the individual participants in the form of booklets. There were two example problems in each booklet: one of them with valid a conclusion (in figure 1 for half of the participants and in figure 2 for the other half of the participants), and one of them with no valid conclusion (in figure 3 for half of the participants and in figure 4 for the other half of the participants). The instructions, which the participants read as a group, were as follows (translated from the Spanish):

This is a test of reasoning. You will be given some problems, and you should respond to each of them. The problems consist of two sentences, and will be similar to this one:

> If Rosa is participating in the experiment, then she is a student.
> If Rosa is a student, then she is young.

As you can see, there is a constituent repeated in both sentences (a student). The conclusion should link the two constituents that are not repeated, and the repeated constituent should not appear in the conclusion. For example, there are two possible conclusions for the previous problem:

> If Rosa is participating in the experiment, then she is young.
> If Rosa is young, then she is participating in the experiment.

You should draw only one of them—the one that you think is correct. But in case you think that neither conclusion can be correctly inferred from the premises, you should select:

> No valid conclusion.

The response format is as follows:

Conclusion:
IF...   THEN...
No valid conclusion

If you think that there is a valid conclusion, fill in the appropriate completions. Instead, if you think that no conclusion follows logically from the premises, then mark the option "No valid conclusion". You could choose this form to answer the previous problem. Now you should try the next example. If you have any doubt about it, please ask the experimenter:

If Rosa is participating in the experiment, then she is a student.
If Rose is young, then she is a student.

Conclusion:
IF...   THEN...
No valid conclusion

If you don't have any questions, please start with the problem on the next page.

The participants then worked their way through the booklet of problems, and they were allowed as much time as they needed to complete the task.

Result

Table 1 presents the percentages of four sorts of response (*If A–C*, *If C–A No valid conclusion* and "other" responses) depending on the structure of the double conditional and on which of the two conclusions was believable. Overall, for the determinate problems, the participants drew 50% valid conclusions, and there was no difference in their percentages for the two figures. However, as predicted, the participants were much more likely to draw the valid conclusion when it was believable (86%) than when it was unbelievable (14%), and the difference was highly reliable (Wilcoxon test, $z = 4.70$, $P < .0001$, one-tail). The effect, contrary to our prediction, did not differ between the two figures. When the valid conclusion was unbelievable, and the converse conclusion believable, the participants tended to respond "no valid conclusion" (50%) rather than to draw the invalid conclusion (31%), although this difference was of marginal significance (Wilcoxon test, $z = 1.41$, $P < .16$, two-tail). For the indeterminate problems, which have no valid conclusion, the participants were more likely to draw the believable conclusion (66%) than to make either of the other two responses (24%; the converse unbelievable conclusion and *no valid conclusion*), and the difference was reliable (Wilcoxon test, $z = 3.28$, $P < .001$, one-tail). In addition, there was a reliable bias favouring A–C conclusions (41% overall) over C–A conclusions (30% overall; Wilcoxon test, $z = 2.20$, $P < .03$; two tail).

TABLE 1
Experiment 1

The Believable	Determinate figures				Indeterminate figures			
	1. If A–B If B–C		2. If B–A If C–B		3. If A–B If C–B		4. If B–A If B–C	
Conclusion:	If A–C	If C–A	If A–C	If C–A	If A–C	If C–A	If A-C	If C–A
Responses:								
∴ If A–C	84	19	38	3	69	3	75	16
∴ If C–A	3	25	9	88	0	63	0	56
No valid conclusion	6	53	47	6	22	25	13	19
Other responses	7	3	6	3	9	9	12	9

The percentages of the three sorts of response in Experiment 1 as function of the figure of the double conditional and of which of the two conclusions was believable. The "other" responses were typically conditional conclusions that violated the instructions by including the middle proposition, B, in place of one of the end propositions.

Discussion

The results corroborated our main predictions. Reasoners tend to draw valid conclusions when they are believable, but to refrain from them when they are unbelievable. Likewise, with double conditionals that have no valid conclusions, reasoners tend to draw believable conclusions rather than to draw the converse conclusions or to respond that there is no valid conclusion. Contrary to our prediction, we failed to observe any difference in performance between the two determinate figures. There were, however, two unexpected biases. First, when a valid conclusion was unbelievable, participants were more likely to respond that there was no valid conclusion than to draw the converse believable conclusion. Second, with the indeterminate figures, there was a small but reliable bias towards If A–C conclusions rather than If C-A conclusions.

The results are readily accounted for by the theory of mental models. It predicts that reasoners are likely to draw valid conclusions for the determinate problems, but where the valid conclusion is unbelievable, they are likely to think again. They can either consider the converse conclusion, which is invalid though believable, or they can eliminate the explicit model and respond that "nothing follows". One possible explanation for the bias towards the latter response is that reasoners are unlikely to draw a converse conclusion in the determinate figures, because it is contrary to the figural bias, and the converse premises, which would render the conclusion valid, are relatively unbelievable. For example, the converse of the believable premise:

If Concha studies, then she passes the exam.

is the less believable assertion:

>   If Concha passes the exam, then she studies.

With indeterminate figures, however, it is easier to construct the converse model because there is no strong figural bias, and so participants will tend to draw whichever of the two conclusions is believable.

Only the failure to detect any difference between the two determinate figures remains to be explained. One hypothesis is that the experimental procedure allowed participants too much time for such a difference to be detected. They have to re-order the premises, i.e. in effect, to convert figure 2 to figure 1, and they can do so merely by starting the process of constructing mental models with the second premise rather than the first. This step is likely to have only a small effect on accuracy in double conditionals, but it may show up in latencies. In the next experiment, we accordingly increased the memory load imposed by the task and timed the latencies of the participants' responses.

EXPERIMENT 2

In order to increase the load on working memory, we used a computer to present the premises constituent by constituent. The participants read a constituent, pressed the space bar, which led to the presentation of the next constituent, and so on. They could not go back to the previous constituents, and so they had to remember them. Such a procedure should increase the difficulty of the task, which might then be sensitive enough to detect a difference between the two determinate figures. Hence, we predicted the same general pattern of results as in the previous experiment, but, in addition, there should be faster latencies for figure 1 than for figure 2, which calls for the premises to be re-ordered in order to draw a valid conclusion.

Method

Design.   The design was the same as for Experiment 1 with the participants carrying out the eight different sorts of problem, which were presented in a different random order to each participant.

Participants.   The participants were 33 first-year psychology students at the University of La Laguna, who received course credit for their participation.

Materials.   The materials and examples were the same as those for Experiment 1. We carried out a post-experimental check of the believability of the respective premises and conclusions. A week after the experimental session, the participants were given the same premises together with the two possible conditional conclusions. Their task was now to evaluate the believability of each

of the four sentences (the two premises and the two possible conclusions) in each of the problems. They judged each sentence as either believable or unbelievable. Because all of the premises but only half of the possible conclusions were believable *a priori*, we added 10 filler conditionals that were also unbelievable in order to avoid a bias towards affirmative responses. The order of the 42 conditional sentences was randomised for this test. Overall, the participants judged 87% of the premises as believable, 84% of the *a priori* believable conclusions as believable, but only 12% of the *a priori* unbelievable conclusions as believable.

Procedure.   The participants were tested individually in a quiet room. The instructions were similar to those of the previous experiment but they explained the use of the computer. The participants drew their own conclusions in their own words, which they spoke aloud. They were given an example of a pair of premises, and told that one proposition was common to both premises. They were asked to draw a conclusion that avoided this proposition. The instructions also explained that there were two possible forms of conclusion (*If A then C* and *If C then A*), and that the participants should draw "only the conclusion that you think could be correctly deduced". If the participants thought that there was no conclusion that followed from the premises, they were to say: "*Sin conclusion valida*" (No valid conclusion).

The materials were presented on the visual display of a computer. At the beginning of each trial, there was a mask in the centre of the screen. When the participant pressed the space-bar on the keyboard, one segment of a conditional premise became visible, and thereafter the next segment became visible as the previous segment was masked. The effect was accordingly of viewing each conditional, constituent by constituent, through a moving window. The second conditional was presented in the same way, but underneath the position of the first conditional. In the following example, the slashes separate the four constituents into which each conditional was divided. These constituents were:

1. "If" and the name of an individual
2. the rest of the antecedent of the conditional
3. "then"
4. the rest of the consequent of the conditional.

For example:

/Si Pedro / enciende la estufa / entonces / se quita el abrigo.
/Si Pedro / tiene frio / entonces / enciende la estufa.
[/If Pedro / turns on the heater/ then/ he takes off his coat.
/If Pedro / is cold/ then / he turns on the heater.]

In Spanish, it is not necessary to include the pronoun ("he" or "she" in the consequent. When the participants pressed the space bar after they had read the last constituent of the second conditional, they were prompted with:

CONCLUSION?

presented in the centre of the screen. They then had to speak their conclusion aloud. The participants had to formulate their conclusions before they began to speak. A voice key activated the computer so that it recorded the latency of the participant's response, i.e. the interval from the onset of the prompt word (CONCLUSION?) to the start of the participant's utterance. The experimenter wrote down the participant's conclusion. All of the possible conclusions started with the same sound: "*Si*" (the conditional "if"), or "*Sin*" (for the "no valid conclusion" response) in order to avoid the effect of a different sensitivity of the voice key to different phonemes.

Results

Table 2 presents the percentages of the four sorts of response (*If A–C*, *If C–A*, *No valid conclusion*, and "other" responses). These responses are shown as a function of the figure of the double conditional and of which of the two conclusions was believable. The table also presents the mean latencies of response to each of the eight sorts of problem. Overall, for the determinate problems, the participants drew 33% of valid conclusions. Unlike Experiment 1, but as we predicted, the percentage of valid conclusions was reliably higher for figure 1 (42%) than for figure 2 (23%; Wilcoxon test, $z = 2.32$, $P < .02$; one tail). As we also predicted, the participants were much more likely to draw a valid conclusion when it was believable (50%) than when it was unbelievable (15%; Wilcoxon test, $z = 3.48$, $P < .0003$, one tail). The effect, again unlike the previous experiment, was now in the direction predicted by the hypothesis. However, the effect was not reliable: in figure 1, the difference was 42%, whereas in figure 2, the difference was only 27% (Wilcoxon test, $z = 0.89$, $P < .19$, one tail). When the valid conclusion was unbelievable, the participants tended to respond "no valid conclusion" (17%) rather than to draw the invalid conclusion (6%). But, as in Experiment 1, this effect was not quite significant (Wilcoxon test, $z = 1.6$, $P < .11$, two-tail). For the indeterminate figures, the participants were more likely to draw a believable conclusion (30%) than to make either of the two other responses (15%, Wilcoxon test, $z = 2.43$, $P < .008$, one-tail). In these figures, there was also a reliable bias favouring *If A–C* conclusions (28%) over *If C–A* conclusions (11%; Wilcoxon test, $z = 2.81$, $P < .005$; two tail).

The difficulty of the task is clear from the high percentage of "other" responses. These were mainly conclusions that contained the middle pro-

TABLE 2
Experiment 2

	Determinate figures				Indeterminate figures			
	1.		*2.*		*3.*		*4.*	
	If A–B		*If B–A*		*If A–B*		*If B–A*	
	If B–C		*If C–B*		*If C–B*		*If B–C*	
The Believable Conclusion:	*If A–C*	*If C–A*	*If A–C*	*If C–A*	*If A–C*	*If C–A*	*If A–C*	*If C–A*
Responses:								
∴ If A–C	64	21	9	0	45	3	42	21
∴ If C–A	3	3	9	36	3	8	9	12
No valid conclusion	3	15	18	6	3	15	3	3
Other responses	30	61	64	58	49	64	46	64
Response latencies (s):	2.1	2.9	3.9	3.2	2.6	3.1	1.7	2.4

The percentages of the four sorts of response in Experiment 2 and the mean latencies of *all* the responses. The data are shown as a function of the figure of the double conditional and of which of the two conclusions was believable. The "other" responses were typically conditional conclusions that violated the instructions by including the middle proposition, B, in place of one of the end propositions.

position, B, instead of one of the end propositions, i.e. they merely repeated one of the premises or its converse. Some of these conclusions, however, included implicit or explicit negations. Thus, for example, participants would negate a valid but unbelievable conclusion, e.g. "If Pedro is cold, then he takes off his coat", in order to render it more believable, e.g. "Pedro is *hot*, then he takes off his coat", or "If Pedro is cold, then he puts *on* his coat" (our italics). Such modifications make better pragmatic sense, and they have been reported by other investigators in reasoning tasks (see e.g. Wason & Johnson-Laird, 1972).

The response latencies showed no overall effect of believability, but the unequal frequencies of response made it impossible to test whether believable valid conclusions were drawn faster than unbelievable valid conclusions. However, responses of whatever sort were faster for the two determinate figures when the believable conclusion was valid. (2.1s + 3.2s) than when the believable conclusion was invalid (2.9s + 3.9s; Wilcoxon test, $z = 1.89$, $P < .03$; one tail). Responses were also faster to figure 1 than to figure 2 (Wilcoxon test, $z = 2.80$, $P < .003$; one tail). For the indeterminate problems, responses were faster to figure 4 than to figure 3 (Wilcoxon test $z = 2.58$, $P < .01$; two tail). They were also faster to draw believable A–C conclusions than believable C–A conclusions although this difference was not quite reliable (Wilcoxon test, $z = 1.58$, $P = .11$; two tail).

Discussion

The most obvious effect of presenting the double conditionals piecemeal was to increase the difficulty of the task. This difficulty was reflected in three ways: (1) the small proportion of valid conclusions to the determinate figures; (2) the small proportion of correct responses to the figures, i.e. "no valid conclusion"; and (3) the large proportion of erroneous responses (the "other" responses) except to the easiest problem (see the first column of Table 2). These "other" responses were nearly always believable, however. The other salient feature of the results was that they corroborated the predictions of the model theory. Thus, reasoners tended to draw valid conclusions when they were believable, but not when they were unbelievable. In the latter case, most participants responded *no valid conclusion*, although the proportion of these responses was smaller than in the previous experiment—perhaps because there were so many "other" responses. With the indeterminate figures, as before, participants tend to draw believable rather than unbelievable conclusions.

One feature of the present experiment was that it successfully demonstrated that figure 1 is easier than figure 2: participants drew more correct conclusions to figure 1 and, most importantly, they responded more rapidly. These results corroborate the hypothesis is that figure 2 calls for the premises to be mentally re-ordered in the process of drawing a conclusion. Likewise, the results also tended to confirm our hypothesis that the bias towards drawing believable conclusions should be stronger for figure 1 than for figure 2. In general, the latencies revealed the same pattern of results as the accuracy of the responses. Thus, participants were faster to response when a valid conclusion was believable. They were also faster to draw *If A–C* conclusions in the indeterminate figures. Only one result was unpredicted; the participants were faster to respond to double conditionals in figure 4 than in figure 3. We have no explanation for this difference.

EXPERIMENT 3

Although the previous experiment corroborated our predictions, the task was very difficult, as evinced by the many "other" responses. The present experiment was therefore designed to ease the load on working memory while still allowing us to measure the latencies of responses.

Method

Design.   The design was the same as in the previous experiments.

Participants.   The participants were 48 undergraduate psychology students at the University of La Laguna who received course credit for their participation. They had not participated in any previous experiment on logical reasoning.

Materials.   We devised eight different sets of contents in the form of double conditionals, and in the experiment as a whole these contents were rotated over the eight conditions (the four sorts of figure and the two believability conditions). The contents were accordingly selected so that each of them could be used to construct problems in all eight conditions. Table 3 illustrates how a single content is used in this way. The crux is that each conditional premise and its converse was believable, but one of the two possible conclusions was believable and one of them was unbelievable. The eight contents concerned simple temporal and causal relations, and each pair of conditionals referred to the same individual—a female for half the materials, and a male for the other half.

TABLE 3
Experiment 3: Contents

Figure 1: If A–B If B–C, with If A–C believable:
  If Pablo turns on the television, then he is at home.
  If Pablo is at home, then he watches the news.

Figure 1: If A–B If B–C, with If C–A believable:
  If Pablo watches the news, then he is at home.
  If Pablo is at home, then he turns on the television.

Figure 2: If B–A If C–B, with If A–C believable:
  If Pablo is at home, then he turns on the television.
  If Pablo watches the news, then he is at home.

Figure 2: If B–A If C–B, with If C–A believable:
  If Pablo is at home, then he watches the news.
  If Pablo turns on the television, then he is at home.

Figure 3: If A–B If C–B, with If A–C believable:
  If Pablo turns on the television, then he is at home.
  If Pablo watches the news, then he is at home.

Figure 3: If A–B If C–B, with If C–A believable:
  If Pablo watches the news, then he is at home.
  If Pablo turns on the television, then he is at home.

Figure 4: If B–A If B–C, with If A–C believable:
  If Pablo is at home, then he turns on the television.
  If Pablo is at home, then he watches the news.

Figure 4: If B–A If B–C, with If C–A believable:
  If Pablo is at home, then he watches the news.
  If Pablo is at home, then he turns on the television.

An example of how the same contents were used in all eight conditions of Experiment 3.

Procedure.   The problems were presented on the visual display of a computer, and were in a different random order for each participant. The participants pressed the space bar, which led to the presentation of the two conditional premises one above the other. When they had formulated their conclusion, they pressed the space bar again, which caused the problem to disappear from the screen, and the computer prompted them with the instruction:

Type the conclusion>

The participant then typed the conclusion and pressed the enter key, which led to the presentation of the instruction:

When ready, press the space bar to continue>

The computer recorded the time from the onset of the premises to the participants' response of pressing the space bar to indicate that they were ready to type their response.

The instructions, as before, stressed the importance of drawing a conclusion that followed from the premises, or of responding *no valid conclusion*. The participants were told to formulate their conclusion completely before they responded, and they were also told to respond as quickly as possible provided that their response was correct. The participants were given two practice problems, one determinate and one indeterminate, prior to the start of the experiment proper.

Results

Table 4 presents the percentages of the four sorts of response (*If A–C*, *If C–A*, *No valid conclusion*, and "other" responses) for each of the eight sorts of problem. The "other" responses, as before, were typically conclusions in which the middle proposition, B, occurred erroneously in place of one of the end propositions. In the present experiment, this percentage was small, as in Experiment 1. The table also presents the mean latencies of the *correct* responses to each of sort of problem. We have presented the means of the correct responses in this experiment because correct responses occurred more often than in the previous experiment. Overall, for the determinate problems, the participants drew 52% valid conclusions. The percentage of valid conclusions was slightly higher for figure 1 (57%) than for figure 2 (47%), but the difference was not quite significant (Wilcoxon test, $z = 1.49$, $P < .07$, one tail). Once again, the participants were more likely to draw a valid conclusion when it was believable (73%) than when it was unbelievable (31%, Wilcoxon test, $z = 4.12$,

$P < .0001$, one tail). However, as in Experiment 1, the effect, contrary to our prediction, did not differ for the two figures: in figure 1, the difference was 40%, whereas in figure 2, the difference was 44% (Wilcoxon test, $z = 0.39$, n.s.). Unlike the two previous experiments, however, when the valid conclusion was unbelievable, the participants were more likely to draw the converse believable conclusion (52%) than to respond "no valid conclusion" (14%; Wilcoxon test, $z = 3.7$, P .0002, two-tail). For the indeterminate figures, the participants were more likely to draw a believable conclusion (57%) than to make either other of the two responses (36%, Wilcoxon test, $z = 2.18$, $P < .02$, one-tail). There was once again a reliable bias favouring If A–C conclusions (38%) over If C–A conclusions (27%; Wilcoxon test, $z = 2.08$, P .04; two tail).

The latencies of the correct responses showed that participants were faster to draw a valid conclusion when it was believable (14.3s + 19.Ss) than when it was unbelievable (19.4s + 23.5s; Wilcoxon test, $z = 2.01$, $P < .03$, one tail). They were also faster to figure 1 problems (14.3s + 19.4s) than to figure 2 problems (23.5s + 19.5s; Wilcoxon test, $z = 2.66$, $P < .004$, one tail). There were no reliable differences for the correct *no valid conclusion* responses to figures 3 and 4.

TABLE 4
Experiment 3: Responses

	Determinate figures				Indeterminate figures			
	1.		2.		3.		4.	
	If A–B		If B–A		If A–B		If B–A	
	If B–C		If C–B		If C–B		If B–C	
The Believable Conclusion:	If A–C	If C–A	If A–C	If C–A	If A–C	If C–A	If A–C	If C–A
Responses:								
∴ If A–C	77	38	52	8	58	15	67	10
∴ If C–A	6	52	25	69	2	52	4	50
No valid conclusion	10	6	21	15	33	23	27	31
Other responses	7	4	2	8	7	10	2	9
Correct response latencies (s)	14.3	19.4	23.5	19.5	25.4	20.3	19.3	23.4

The percentages of the three sorts of response in Experiment 3 and the mean latencies of the *correct* responses. The data are shown as a function of the figure of the double conditional and of which of the two conclusions was believable. The "other" responses were typically conditional conclusions that violated the instructions by including the middle proposition, B, in place of one of the end propositions.

Discussion

In general, the results again corroborated the predictions of the model theory. With determinate figures, the participants drew valid conclusions more often when they were believable than when they were unbelievable. With in-determinate figures, they tended to draw whichever invalid conclusion was believable. The reported latencies in the present experiment were only for the correct responses. They were much longer than those of Experiment 2 because they included the time to read the premises. Nevertheless, they showed that participants drew valid conclusions more quickly when they were believable than when they were unbelievable, and when the problems were in figure 1 than when they were in figure 2. One striking finding was that when a valid con-clusion was unbelievable, the participants in the present experiment, unlike those in the two previous experiments, tended to draw the believable but invalid conclusion rather than to respond "no valid conclusion".

GENERAL DISCUSSION

Our experiments have revealed several robust phenomena about reasoning with double conditionals. With determinate problems in figure 1:

> If A then B
> If B then C

and in figure 2:

> If B then A
> If C then B

participants draw a valid conclusion (*If A then C* for figure 1, and *If C then A* for figure 2) more often when it is believable than when it is unbelievable. This effect is highly reliable: it occurred in all three experiments, and, moreover, the participants in Experiments 2 and 3 drew the believable valid conclusions more quickly than the unbelievable valid conclusions. The model theory predicts this effect. Given a double conditional, such as the premises in figure 1, participants should construct the following initial models:

> a   b   c
>   …

If the explicit model is believable, then the process goes smoothly; but if the explicit model is unbelievable, then reasoners are likely to reject it. In the first two experiments, they tended to respond that there was no, valid conclusion. Thus, when they were left with only the implicit model just given (denoted by

the ellipsis), they inferred that nothing followed. In Experiment 3, however, they tended instead to draw the believable invalid conclusion. The reason, we believe, has to do with the materials in Experiment 3. In the previous experiments, the conditional premises, such as:

> If Concha studies, then she passes the exam.

have converses that are less believable:

> If Concha passes the exam, then she studies.

In contrast, the conditional premises in Experiment 3 were deliberately devised so that their converses were equally believable, e.g.:

> If Pablo is at home, then he turns on the television.

has the converse:

> If Pablo turns on the television, then he is at home.

Indeed, in this experiment, each conditional and its converse occurred equally often as a premise. Hence, after the participants had abandoned an explicit model, such as:

> a   b   c

because it supported an unbelievable conclusion, they may have been more prepared to construct the converse explicit model:

> c   b   a

because the converse of the conditional premises are believable. The new model supports a believable conclusion, and so the participants draw it even though it is technically invalid.

There was a tendency for participants to be more accurate and to respond more quickly to figure 1 than to figure 2. And this phenomenon is to be expected on the grounds that participants re-order the two premises in figure 2 to construct the models of them (see Johnson-Laird & Bara, 1984).

The effect of believability was particularly striking with the indeterminate problems of figure 3:

> If A then B.
> If C then B.

and figure 4:

> If B then A.
> If B then C.

Strictly speaking, they have no valid conclusion unless at least one of the conditionals is interpreted as a bi-conditional. The model theory, however, predicts that reasoners can construct either sort of initial models for them:

>   a   b   c
>   …

or:

>   c   b   a

because there is no strong figural bias in these figures. Overall, the participants tended to draw whichever of the two conclusions was believable. But the majority of conclusions were of the form: *If A then C*, which suggests that there is a bias towards the former sort of models. A similar bias has been detected in the study of syllogisms (see Johnson-Laird & Byrne, 1991). It may well reflect the "first in, first out" principle of working memory (Broadbent, 1958). With these symmetric figures, reasoners may tend to build models based on the first premise and to abandon them only if the result is unbelievable.

The inferential task was difficult for participants in Experiment 2, presumably because each premise was presented one constituent at a time with no opportunity for the participants to go back to read earlier constituents. As a consequence, a high proportion of their responses were egregious errors in which, for example, they substituted the middle proposition, B, in place of one of the end terms, A or C. Once again, however, there were manifest effects of believability. These erroneous conclusions were almost always believable, and the participants would even on occasion distort the content of the materials in order to negate what would otherwise have been an unbelievable conclusion, e.g. instead of inferring "If Pablo feels sleepy, then he goes to bed late", they erred by drawing the believable conclusion, "If Pablo feels sleepy, then he goes to bed early".

To what extent might our results be accounted for by alternative theories? First, the *misinterpreted necessity* hypothesis cannot explain them (see Dickstein, 1980, 1981; Evans, 1989; Evans et al., 1983). Even if the participants had misinterpreted necessity, they should still have drawn the valid conclusions to the determinate problems regardless of their believability. These conclusions are necessarily true, and the hypothesis proposes that reasoners are affected by

believability only if they fail to infer that a conclusion is necessarily true or necessarily false.

Second, the results cannot be explained by the *selective scrutiny* hypothesis (Evans, 1989; Evans et al., 1983). This hypothesis postulates that reasoners assess the believability of a conclusion, and accept it if it is believable. They only assess whether a conclusion follows validly if it is not believable, and then accept it provided that it is valid. Hence, if selective scrutiny were true, then our participants should have accepted the unbelievable *valid* conclusions in all three experiments. They did not do so, but instead tended to assert that nothing follows (Experiments 1 and 2), or to draw the converse invalid conclusion (Experiment 3).

Finally, if participants were reasoning on the basis of formal rules of inference, then the content of the problems could influence only their initial interpretation of the conditionals or their rejection of unbelievable conclusions after the process of reasoning is complete. Proponents of formal rules have emphasised the importance of pragmatic principles in the interpretation of conditionals (see e.g. Braine & O'Brien, 1991; Rips, 1994). Because all the premises in the experiments were believable, the most relevant idea is that reasoners make an "invited inference" (Geis & Zwicky, 1971) from a premise, e.g. they infer that the converse of the conditional is true too. This step would, of course, justify drawing conclusions from the indeterminate figures—in effect, it turns a conditional into a bi-conditional. However, there are no interpretations of conditionals, and no invited inferences, that justify the response of *no valid conclusion* to the determinate figures. Similarly, it is hard to see how reasoners would draw erroneous conclusions that implicitly negate the correct conclusions if their process of reasoning was solely controlled by syntactic, as opposed to semantic, processes.

Perhaps resourceful defenders of misinterpreted necessity, selective scrutiny, or formal rules, could develop a *post hoc* account of the effects of logical structure and belief that we have demonstrated in our experiments. Meanwhile, the model theory provides a feasible account of them.

REFERENCES

Barston, J.I. (1986). *An investigation into belief biases in reasoning.* Unpublished doctoral dissertation, Plymouth Polytechnic, UK.

Braine, M.D.S., & O'Brien, D.P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, *98*(2), 182–203.

Broadbent, D.E. (1958). *Perception and communication.* New York: Pergamon Press.

Dickstein, L.S. (1978). The effect of figure on syllogistic reasoning. *Memory and Cognition*, *6*, 76–83.

Dickstein, L.S. (1980). Inference errors in deductive reasoning. *Bulletin of the Psychonomic society*, *6*, 414–416.

Dickstein, L.S. (1981). Conversion and possibility in syllogistic reasoning. *Bulletin of the Psychonomic Society*, *18*, 229–232.

Evans, J.St.B.T. (1989). *Bias in human reasoning: causes and consequences*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Evans, J.St.B.T., Barston, J.L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*, 295–306.

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

Evans, J.St.B.T., & Over, D.E. (1996). *Rationality and reasoning*. Hove, UK: Erlbaum (UK) Taylor & Francis Ltd.

García-Madruga, J.A. (1982). Un estudio sobre el efecto de la figura en el razonamiento silogístico. *Estudios de Psicología*, *11*, 23–32.

Geis, M.L., & Zwicky, A.M. (1971). On invited inferences. *Linguistic Inquiry*, *2*, 561–566.

Henle, M., & Michael, M. (1956). The influence of attitudes on syllogistic reasoning. *Journal of Social Psychology*, *44*, 115–127.

Johnson-Laird, P.N., & Bara, B.G. (1984). syllogistic Inference. *Cognition*, *16*, 1–61.

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

Johnson-Laird, P.N., & Savary, F. (1996). Illusory inferences about probabilities. *Acta Psychologica*, *93*, 69–90.

Johnson-Laird, P.N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, *10*, 64–98.

Newstead, S.E., Pollard, P., Evans, J.St.B.T., & Allen, J.L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, *45*, 257–284.

Oakhill, J.V., & Johnson-Laird, P.N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology*, *37*A, 553–69.

Oakhill, J.V., Johnson-Laird, P.N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition, 31,* 117–40.

Revlin, R., & Leirer, V.O. (1978). The effect of personal biases on syllogistic reasoning: Rational decisions from personalized representation. In R. Revlin, & R. Mayer (Eds.), *Human reasoning*, Washington, DC: Winston.

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

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning*. London: Batsford. Cambridge, MA: Harvard University Press.

Wilkins, M.C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology*, *16*, 1–83.