Logic, Models, and Paradoxical Inferences

ISABEL ORENES AND P. N. JOHNSON-LAIRD

Abstract: People reject 'paradoxical' inferences, such as: *Luisa didn't play music; therefore, if Luisa played soccer, then she didn't play music*. For some theorists, they are invalid for everyday conditionals, but valid in logic. The theory of mental models implies that they are valid, but unacceptable because the conclusion refers to a possibility inconsistent with the premise. Hence, individuals should accept them if the conclusions refer only to possibilities consistent with the premises: *Luisa didn't play soccer; therefore, if Luisa played a game then she didn't play soccer*. Two experiments corroborated this prediction for three sorts of 'paradox', including a disjunctive paradox.

1. Introduction

Some sentential connectives have interpretations that are truth functional and other interpretations that are not truth functional. For example, the conjunction:

Pat is tall and Viv is short,

is normally interpreted as truth functional, i.e. it is true if and only if both of its clauses are true, and so it is synonymous with:

Viv is short and Pat is tall.

In contrast, the assertion:

Pat pushed Viv and Viv fell over,

is normally interpreted, not as truth functional, but as implying a temporal order between the two events: the first event occurred before the second event, and perhaps caused it. Hence, the assertion is not synonymous with:

Viv fell over and Pat pushed her.

This research was supported by grant SEJ2006-14714 and Consolider-Ingenio 2010 CSD2008-00048 from the Spanish Government to the first author, and by National Science Foundation Grant SES 0844851 to the second author to study deductive and probabilistic reasoning. We thank David Beltrán, Nick Chater, Jonathan Evans, Vittorio Girotto, Sangeet Khemlani, Klaus Oberauer, Carlos Santamaría, and Walter Schroyens, for their thoughts about the present research. We also thank Mike Oaksford and two anonymous reviewers, for their criticisms of previous drafts of the article.

Address for correspondence: Isabel Orenes, Universidad de La Laguna, Facultad de Psicología, Departamento de Psicología Cognitiva, Social, y Organizacional, Campus Guajara, sn. 38205 La Laguna, Tenerife, Spain.
Email: iorenes@ull.es

Mind & Language, Vol. 27, No. 4 September 2012, pp. 357–377.

One way to account for these differences is to postulate that ‘and’ is ambiguous, and has both a truth-functional and a non-truth-functional interpretation. Grice (1989) made an analogous argument, and he extended his analysis to conditionals, such as:

If Pat pushed Viv then she fell over.

He argued that the meaning of conditionals is truth functional. They are true if their if-clauses are false or their then-clauses are true, or both, and so they are true in every case except when their if-clauses are true and their then-clauses are false. This semantics corresponds to the connective of ‘material implication’ in logic. Table 1 shows its truth table. But, Grice also argued that pragmatic factors could overrule these interpretations. Not all theorists, however, accept this view (e.g. Stalnaker, 1968; Evans and Over, 2004). And they have an apparently cogent reason for rejecting it, which concerns the so-called ‘paradoxes’ of material implication—the topic of the present paper.

The paradoxes of material implication are not real paradoxes, but they are sufficiently perplexing that we will continue to use this term to describe them (see Pfeifer and Kleiter, 2011; Schroyens, 2010). To understand the paradoxes, we must first clarify the meaning of ‘validity’, which is open to more than one interpretation. The usage we follow is standard in modern logic that recognizes the distinction between proof theory and model theory: an inference is ‘valid’ if its conclusion must be true given the truth of its premises, i.e. the conclusion is true in every case in which the premise is true (Jeffrey, 1981, p. 1). As a corollary, an inference is invalid if its conclusion could be false even though its premises are true. Any inference is accordingly either valid or not valid (invalid).

A key question about the paradoxes is whether or not they are valid. There are two sorts of paradox, and the first sort is illustrated here:

Lucia didn’t wear the shoes. (Not-B)
Therefore, If Lucia wore jewelry then she didn’t wear the shoes. (∴ If A then not-B)

The categorical premise asserts the truth of the conditional’s then-clause. As Table 1 shows, a material implication is true in this case regardless of whether its if-clause is true or false. The inference is accordingly valid for material implication, because the truth of the premise suffices for the truth of the conclusion.

Pat is here	Viv is here	If Pat is here then Viv is here
True	True	True
True	False	False
False	True	True
False	False	True

Table 1 *An illustration of the truth table for material implication in logic.*

The second sort of paradox is illustrated here:

It won't rain today.	(Not-A)
Therefore, if it rains today then the forecast is right.	(∴ If A then B)

The categorical premise denies the conditional's if-clause, and, as Table 1 also shows, the inference is valid for material implication. Yet, both sorts of paradox are implausible, and individuals tend to reject them.

On this account, if conditionals are interpreted as material implications then the paradoxes are valid in semantic sense that we have adopted from modern logic: they preserve truth. But, their unacceptability has led some theorists to argue that conditionals in daily life are not interpreted as material implications. Stalnaker (1968) argued that the falsity of the antecedent is never sufficient reason to affirm a conditional, even an indicative conditional, and so the paradoxes are invalid (cf. Stalnaker, 1975). Likewise, Evans and Over (2004, p. 153) wrote: 'For us, the original sin of JLB [Johnson-Laird and Byrne, 2002] theory is their endorsement of the logical validity of the paradoxes of interpreting a natural language conditional as truth functional. We hold that the paradoxes are logically invalid for natural language conditionals, including ''basic'' conditionals'. Hence Stalnaker, Evans and Over, and others, reject the possibility of a truth-functional meaning for 'if'.

In summary, two principal views exist about the paradoxes. One view, which we defended above, is that they are valid, but unacceptable. The other view is that the paradoxes are invalid and unacceptable. The two views naturally arise from different conceptions of the meaning of conditionals. But, they can also arise from a different conception of validity, such as that it is inherently probabilistic (e.g. Adams, 1975). The correct analysis of validity does not appear to be susceptible to empirical inquiry. Hence, as we will show, the crucial result in our research is the discovery of semantic factors that make both sorts conditional paradox, and an analogous disjunctive paradox, acceptable to logically naïve participants. Our approach is based on the theory of mental models (e.g. Johnson-Laird, 1983). Unlike previous accounts (e.g. Bonnefon and Politzer, 2011; Grice, 1989; Rips, 1994), our investigation of the paradoxes focuses on semantics rather than pragmatics. It demonstrates that with certain *contents* individuals do accept the paradoxical inferences as valid, where we emphasize that 'valid' refers to the semantic definition above, which is reflected in the instructions to the participants in our experiments. Their results have implications for current psychological theories of reasoning, and that is why an investigation of the paradoxes is worth making. We consider these other theories later in the paper, but next we describe the theory of mental models, because it is going to do some work for us.

2. Mental models

The theory of mental models—the model theory, for short—postulates that individuals think about the possibilities consistent with the premises and with their knowledge, and infer that a conclusion follows validly if it holds in all these possibilities (Johnson-Laird, 2006; Johnson-Laird and Byrne, 1991). A conditional such as, *If Pat is in Rome then she is in Italy*, has the grammatical form, *If A then B*, and refers to three possibilities, which we also symbolize here:

Pat is in Rome and is in Italy.	(A B)
Pat is not in Rome and is in Italy.	(not-A B)
Pat is not in Rome and is not in Italy.	(not-A not-B)

We use sentences to describe the contents of models in this article, but the theory postulates that in reality individuals use iconic mental models of actual situations—iconic in the sense that the structure of a model corresponds to the structure of what it represents (see Peirce, 1931–1958, Vol. 4, for the concept of iconicity). The three preceding possibilities reflect the *core* meaning of the conditional (Johnson-Laird and Byrne, 2002). They correspond to the cases in Table 1 in which the conditional is true, and so they correspond to material implication, though models represent possibilities instead of truth values. Studies show that participants can list these three possibilities for conditionals (e.g. Barres and Johnson-Laird, 2003; Barrouillet, Grosset and Lecas, 2000). However, everyday reasoning tends to rely on just a single explicit mental model of the conditional in which both its clauses are true, and an implicit model with no specific content to represent the other possibilities in which its if-clause is false (Johnson-Laird and Byrne, 2002). For the conditional above, the mental models are as follows:

Pat is in Rome   in Italy
. . .

The explicit model represents the possibility in which both clauses in the conditional are true, and the implicit model, denoted by the ellipsis, represents the possibilities in which the if-clause is false. When individuals retain all this information, they can flesh out these mental models into fully explicit models of the three possibilities listed above.

The model theory postulates a mechanism of *modulation* in which the meanings of clauses, their referents, and knowledge, can transform the core meanings of connectives, including conditionals, into an indefinite number of other meanings (Johnson-Laird and Byrne, 2002). Modulation is of two sorts. The first sort introduces relations between the events referred to in conditionals, including temporal and other sorts of relation between the events referred to in the conditional's two clauses, as in the case of, say, *if Pat pushed Viv then Viv fell over*. Only the second sort of modulation, however, was manipulated in the present studies. Its effect is to block the construction of models of possibilities from the core meaning of a connective. (The theory postulates that knowledge cannot add models of possibilities, but solely block their construction, see Johnson-Laird and Byrne, 2002). Consider the conditional:

If Pat is in Italy then she is not in Rome.   (If A then not-B)

It refers to only two possibilities

Pat is in Italy and is not in Rome.	(A not-B)
Pat is not in Italy and is not in Rome.	(not-A not-B)

Knowledge blocks the construction of a model of the case: Pat is not in Italy and Pat is in Rome (not-A B), because this case is impossible given that Rome is in Italy. Modulation can block other possibilities in the core interpretation of conditionals to yield different sorts of meaning (see Johnson-Laird and Byrne, 2002).

Experiments have corroborated the occurrence of both sorts of modulation (e.g. Quelhas, Juhos and Johnson-Laird, 2010; Santamaría, García-Madruga and Johnson-Laird, 1998). Although the core meaning corresponds to the truth-function of material implication, modulation yields interpretations of conditionals that are not truth functional. Hence, the system of interpretation as a whole is not purely truth functional (Byrne and Johnson-Laird, 2009; Johnson-Laird and Byrne, 2002). It also establishes, where relevant, temporal or other relations between the events referred to in the if-clause and the then-clause.

How are we to test the model theory of conditionals? One way is to compare its predictions with those of other theories for straightforward inferences, and it has had some success (e.g. Oberauer, 2006; Verschueren, Schaeken and d'Ydewalle, 2005). But, a novel way is to address the paradoxes head on. The model theory explains the reason that individuals usually reject the paradoxes, but it also predicts the sorts of conditionals that should make the paradoxes acceptable. In their grammatical forms in our studies, both paradoxes have a negative premise, and the two conditional paradoxes are, respectively:

1. Not B.
   Therefore, if A then not B.

and:

2. Not A.
   Therefore, if A then B.

The conditional in the first paradox normally refers to three possibilities:

A	not-B
not-A	not-B
not-A	B

Hence, in every case in which the categorical premise holds the conditional conclusion holds, and so the inference is valid (on the semantic account of validity). Naïve individuals list the three possibilities, and yet they reject the inference. This conflict is puzzling, but the model theory solves the puzzle.

An inference is valid if its conclusion holds in all the possibilities in which its premise holds (Jeffrey, 1981); but the converse relation is not required for validity.

According to the model theory, individuals reject the paradox: *not B*; therefore, *if A then not B*, because its premise, *not B*, conflicts with the third of the possibilities above, *not-A* and *B*, to which the conditional refers. In other words, the conclusion throws away semantic information in a radical manner (for the concept of semantic information, see Bar-Hillel and Carnap, 1964): even though the conclusion is valid, it refers to a possibility inconsistent with the premise. Hence, if modulation blocks this third possibility, the inference should be more acceptable. A way in which to block this possibility is to use contents in which the falsity of the if-clause implies the truth of the then-clause. This abstract recipe becomes much clearer with an example. In the following conditional, if the if-clause is false then the then-clause is true:

If Lucia wore jewelry then she didn't wear the bracelet. (if A then not B)

That is, if Lucia didn't wear jewelry then of course she didn't wear a bracelet, because bracelets are jewelry. The conditional therefore refers to only two possibilities:

Lucia wore jewelry and didn't wear the bracelet.	(A not-B)
Lucia didn't wear jewelry and didn't wear the bracelet.	(not-A not-B)

The conflict with the premise no longer occurs: the premise (not-B) matches both possibilities to which the conditional refers, and so individuals should accept the inference:

Lucia didn't wear the bracelet.	(Not B)
Therefore, if Lucia wore jewelry then she didn't wear the bracelet.	(∴ If A then not B)

In sum, the theory makes this first general prediction:

1. *A conditional in which a false if-clause implies a true then-clause should create an acceptable version of the paradox: Not B; ∴ If A then not B.*

This account generalizes to disjunctive inferences of the sort:

Viv is here.	(B)
Therefore, Pat is here or Viv is here, or both.	(∴ A or B)

The disjunctive conclusion is consistent with three possibilities:

Pat is here and Viv is not here.	(A not-B)
Pat is not here and Viv is here.	(not-A B)
Pat is here and Viv is here.	(A B)

The disjunction therefore holds in every case in which the premise holds, and so the inference is valid. Yet, the inference tends to be rejected by logically-untrained individuals (see, e.g., Rips and Conrad, 1983; Schroyens, 2009). The model theory postulates that it is unacceptable because the premise conflicts with the first of the three possibilities to which the conclusion refers: A not-B. Hence, if modulation blocks this possibility, the theory predicts that the inference should tend to be accepted as valid. The way to block this possibility is to use contents in which if *A* is true then B is true, e.g.:

Lucia wore the bracelet or she wore jewelry. (A or B)

This disjunction refers to only two possibilities:

Lucia wore the bracelet and she wore jewelry. (A B)
Lucia didn't wear the bracelet and she wore jewelry. (not-A B)

Once again, the knowledge that a bracelet is jewelry blocks the possibility in which Lucia wore the bracelet and didn't wear jewelry (A not-B). The premise is consistent with both possibilities to which the disjunction refers, and so the inference should tend to be accepted:

Lucia wore jewelry.
Therefore, Lucia wore the bracelet or she wore jewelry.

The theory accordingly makes this second general prediction:

2. A disjunction in which a true first clause implies a true second clause should create an acceptable version of the paradox: B; ∴ A or B, or both.

The second sort of conditional paradox is more egregious:

Not A.
Therefore, if A then B.

The conditional, as we have seen, normally refers to three possibilities:

A	B
not-A	B
not-A	not-B

The premise conflicts with the first of these possibilities, which is also the only possibility to be represented in the conditional's explicit mental model. To block its construction, or at least to render it less likely, modulation needs to yield a so-called 'tollens' interpretation of the conditional (see Johnson-Laird and Byrne, 2002, p. 662). This interpretation is elicited by certain idioms, such as, *if it rains today then I'll eat my hat*, which refers to just one possibility: it won't rain today and I won't eat my hat. But, in order to compare the inference with the other sorts of paradox, we took pains to avoid idioms. Consider instead a conditional, such as:

If it rains today then the forecast is wrong.

It has an interpretation similar to the idiom. That is, modulation implies that the forecast was that it won't rain today, and so if it does rain today the forecast is wrong. The content of the forecast is therefore identical to the premise: it won't rain today. In this context, the conditional can be construed as:

> If it rains today, which it won't, then the forecast is wrong, which it isn't.

This construal corresponds to the tollens interpretation, which refers to just one possibility:

> It won't rain today and the forecast isn't wrong. (not-A not-B)

Hence, modulation has blocked the construction of the possibility that normally conflicts with the premise, and so the inference should tend to be accepted:

> It won't rain today.
> Therefore, if it rains today then the forecast is wrong.

The theory accordingly makes this third general prediction:

> *3. A conditional in which a true then-clause implies a false if-clause should create an acceptable version of the paradox: Not-A; ∴ If A then B.*

In what follows, we report two experiments that test the model theory's predictions about the paradoxes. They do so by comparing the original sort of paradoxes with inferences of the same grammatical form but in which modulation creates a conditional conclusion referring to possibilities that no longer conflict with the premise. The manipulation is a semantic one, because one or two words are substituted for others in the conclusions, and the substitutions create the semantic relations between the two clauses called for in principles 1–3 above. The experiments examined the two conditional paradoxes and the disjunctive paradox. Their transformation into acceptable inferences calls for modulation to yield conclusions that eliminate any offending possibility, i.e. one to which the conclusion refers but that is inconsistent with the premise. The model theory will be corroborated if individuals accept these modulated versions of the inferences. The idea underlying the study is accordingly akin, say, to a study of the Müller-Lyer illusion in which a change to the arrow heads reduces the illusion in a way that corroborates one theory of the illusion, but stands in need of explanation by alternative theories (see, e.g., Gregory, 1997).

3. Experiment 1

The participants evaluated whether or not conclusions of the two sorts of conditional paradox (*Not B*; therefore, *If A then not B*; and *Not A*; therefore, *If A then B*) followed from the premises. The experiment manipulated the contents of assertions so that conditionals in the paradoxical case had the core interpretation referring to a possibility inconsistent with the premise, but conditionals in the modulated case no longer referred to this possibility.

Method

Participants

The participants were 22 undergraduates at La Laguna University, Tenerife (Spain), who carried out the experiment as a course requirement.

Design

The participants evaluated whether or not the conditional conclusions for the two sorts of paradox followed from the premises. There were two sorts of content for each of the two sorts of inference: paradoxical contents that should inhibit the inferences, and matched modulated contents, differing by only one or two words, that should elicit the inferences. The participants carried out five inferences for each of the four sorts (two forms x two contents), yielding a total of 20 inferences, which were presented to each participant in a different random order.

Materials

To create the contents, we used the simple procedures (described informally in the previous section) to devise matching pairs of modulated materials that should elicit the inference and paradoxical materials that should inhibit the inference. The procedure for the first sort of paradox was to devise conditionals of the grammatical form, *if A then not B*, in which the falsity of the if-clause (*not A*) implies the truth of the then-clause (*not B*), e.g.: *if Andrés played a game then he didn't play soccer*. This modulated conditional should yield an inference that should tend to be accepted as valid. The paradoxical version of the inference was then constructed by changing one or two words so that *not A* no longer implied *not B*, e.g. *if Luisa played a game then she didn't play music*. For the second sort of paradox, the procedure was to devise conditionals of the grammatical form, *if A then B*, in which the then-clause implied the falsity of the if-clause, e.g. *if the experiment works then this prediction is false*, where 'this prediction' refers in context to a prediction that the experiment won't work. This modulated conditional should yield an acceptable inference. The paradoxical version of the inference was constructed by changing one or two words so that the then-clause no longer implied the falsity of the if-clause, e.g. *if the experiment works then this prediction is true*. The full materials translated from the Spanish are at: http://mentalmodels.princeton.edu/projects/paradoxes/ (and also in Appendix A). We considered the use of filler materials, but no simple categorical premises lead to obvious cases of valid, or invalid, inferences of conditional conclusions; and to introduce a different sort of inference into the experiment did not seem necessary.

Procedure

The participants were told that the experiment concerned reasoning, but that it was not a test of intelligence. The key instructions translated from the Spanish were:

'On each trial, you'll be asked whether a conclusion follows from an assertion, that is, given that the assertion is true, is it the case that the conclusion must be true? You should answer the question, either "yes" or "no".' The participants circled "Yes" or "No" to make their responses. They were given a simple practice inference based on a conjunction, and then carried out the experiment proper. A typical paradoxical problem had the following format:

> It won't rain today.
> Does it follow that if it does rain today then the forecast is right?

The analogous modulated version of the problem was:

> It won't rain today.
> Does it follow that if it does rain today then the forecast is wrong?

Each problem was presented on a separate sheet of paper, and the participants were allowed to take as much time as they wanted to make their response.

Results

The percentages of 'yes' responses, indicating that the participants accepted the inferences as valid, were as follows:

Not B. Therefore, If A then not B.	Modulated contents:	60%
	Paradoxical contents:	24%
Not A. Therefore, If A then B.	Modulated contents:	74%
	Paradoxical contents:	5%

The participants accepted more inferences with the modulated contents than with the paradoxical contents for both sorts of inference (Wilcoxon tests, $z = 3.09$, $p < .0025$; and $z = 4.06$, $p < .00005$, one-tail tests, respectively). A by-materials analysis corroborated these results (Wilcoxon tests, $z = 2.03$, $p < .05$; and $z = 2.03$, $p < .05$, respectively). As the percentages show, the difference between the two contents tended to be larger for the second sort of inference than for the first, and this interaction was reliable (Wilcoxon test, $z = 2.77$, $p < .01$, two-tail test). The model theory accounts for this tendency. For the first sort of inference, the premise matches the most salient mental model of the conditional (*A and B*), but for the second sort of inference, it does not. Hence, the second sort of paradox is more egregious than the first, and the remedial effects are correspondingly greater.

4. Experiment 2

The experiment examined the disjunctive paradox: *B*; therefore, *A or B*, and an analogous conditional paradox. In both cases, it contrasted a paradoxical version that should inhibit the inference with a modulated version that should elicit the inference. A paradoxical version of the disjunctive inference was, e.g.:

> Eva read a newspaper.

Does it follow that Eva read Don Quixote or she read a newspaper?

The corresponding modulated version was:

Ana read a novel.
Does it follow that Ana read Don Quixote or she read a novel?

Method

Participants

The participants were 20 undergraduates from the same population as before.

Design

There were two sorts of inference, one with a disjunctive conclusion: *B; therefore, A or B*; and one with a conditional conclusion: *Not-B; therefore, if A then not-B*. Each sort of inference had two sorts of contents: paradoxical contents and modulated contents. The participants acted as their own controls and carried out five inferences for each of the four sorts, yielding a total of 20 inferences, which were presented to each participant in a different random order.

Materials and Procedure

The conditional paradoxes were constructed in the same way as before. The disjunctive paradoxes were constructed so that for the modulated versions, which should be acceptable: *A* implied *B* in the disjunction, *A or B*, e.g. *Ana read Don Quixote or she read a novel*. The paradoxical versions, which should be unacceptable, were constructed by changing one or two words, so that *A* no longer implied *B*, e.g. *Eva read Don Quixote or she read a newspaper*. The full materials are at: http://mentalmodels.princeton.edu/projects/paradoxes/ (and also in Appendix B). The instructions and procedure were the same as in Experiment 1.

Results

The percentages of inferences that the participants made were as follows:

B. Therefore, A or B, or both.	Modulated contents:	79%
	Paradoxical contents:	29%
Not-B. Therefore If A then not B.	Modulated contents:	63%
	Paradoxical contents:	7%

The participants accepted more inferences with the modulated contents than with the paradoxical contents for the disjunctions (Wilcoxon test, $z < 3.25$, $p = .001$) and for the conditionals (Wilcoxon test, $z = 3.64$, $p < .001$). A by-materials analysis corroborated these results (Wilcoxon tests, $z = 2.02$, $p < .05$; and $z = 2.03$, $p < .05$, respectively). There was no reliable difference in the effect for the two sorts of paradox (Wilcoxon test, $z = .281$, $p > .5$, two-tail test).

5. General Discussion

The results of both experiments corroborated the model theory. In Experiment 1, participants tended to reject the first sort of paradox, such as:

> Lucia didn't wear the shoes.
> Therefore, if Lucia wore jewelry then she didn't wear the shoes.

But, they tended to accept inferences with modulated conditionals, such as:

> Lucia didn't wear the bracelet.
> Therefore, if Lucia wore jewelry then she didn't wear the bracelet.

In this latter case, the meanings of *jewelry* and *bracelet* modulate the interpretation of the conditional to block a possibility (Lucia didn't wear jewelry but did wear the bracelet) that would otherwise conflict with the premise. No such modulation occurs in the paradoxical case, and the conflicting possibility to which the conditional refers (Lucia didn't wear jewelry and did wear the shoes) leads individuals to reject the inference.

Experiment 1 yielded similar results for the second sort of paradox. The participants tended to reject a paradoxical inference, such as:

> Lorraine will win her tennis match.
> Therefore, if Lorraine doesn't win her tennis match then this expectation is correct.

In the interpretation of the conditional, 'this expectation' is likely to refer to the case that Lorraine won't win, and this possibility conflicts with the premise. But, when the inference was instead:

> Lorraine will win her tennis match.
> Therefore, if Lorraine doesn't win her tennis match then this expectation is incorrect.

The expectation now is that Lorraine will win, which implies the falsity of the if-clause. Its falsity matches the premise, which no longer conflicts with the otherwise salient possibility to which the conditional refers. Participants accordingly tended to accept the inference.

The manipulations in our experiments are not necessarily the only way to make the inferences acceptable. An inference of the same sort as the first paradox should also be acceptable on pragmatic grounds:

> There's ten dollars in the teapot.
> Therefore, if you need some money then there's ten dollars in the teapot.

The conditional has a 'relevance' interpretation according to Johnson-Laird and Byrne (2002, p. 651): it states as a matter of fact that ten dollars is in the teapot, and its if-clause merely refers to a situation in which the fact may be relevant. An experiment has corroborated the occurrence of such interpretation—see Experiment 3 in Quelhas, Johnson-Laird and Juhos (2010). A pretest to this experiment showed, as predicted, that the majority of responses (87%) listed only possibilities in which the then-clause of the conditional held, and the experiment itself showed, again as predicted, that a new group of participants accepted a modus ponens inference from such conditionals but not a modus tollens inference. For conditionals with a core logical interpretation, they accepted both modus ponens and modus tollens inferences at almost equal and high rates (over 90%). Bonnefon and Politzer (2011) make an analogous prediction about the first sort of paradox on pragmatic grounds, but they do not consider the second sort of paradox. They argue that the first sort should be acceptable if the *relevance* of the then-clause depends on the truth of the if-clause, and the speaker can reasonably be expected not to be in a position to know that the if-clause is false. We accept that pragmatics can lead to the acceptability of the inference (see also Sperber and Wilson, 1995), but the manipulation in our experiments was semantic. Bonnefon and Politzer argue that their account makes more accurate predictions than their construal of the model theory, but they present no experimental results corroborating their intuitions.

Certain idioms make the second sort of paradox acceptable. When the second author asserts, as is his wont:

> If that experiment works then I'll jump in Lake Carnegie,

he means that the experiment won't work. Hence, the inference from the denial of the if-clause should be acceptable. Likewise, a then-clause that is patently false should yield the tollens interpretation and an acceptable inference, as in the following example, which we owe to Klaus Oberauer (personal communication, September 2010):

> The experiment will work.
> Therefore, if the experiment doesn't work then London is in Greenland.

We considered using conditionals with then-clauses that were obviously false, as in this example, but decided that their oddity might confuse naïve participants.

Both Rips's (1994, p. 47) theory and the model theory (Johnson-Laird and Byrne, 1991, p. 74) predict that similar paradoxes should occur with inferences yielding disjunctive conclusions, such as:

> David visited England.
> Therefore, David visited Paris or he visited England.

The conclusion refers to a possibility inconsistent with the premise (David visited Paris and he didn't visit England), and so, as Experiment 2 showed, most participants rejected the inference. But, once again, when modulation blocked the offending possibility, most participants accepted the inference, e.g.:

> Paco visited France.
> Therefore, Paco visited Paris or he visited France.

In the course of our research, skeptics have tried to refute our results, and their claims fall into two main categories—those that suggest that are experiments are faulty, and those that suggest that the underlying logic of our research is wrong. We consider both sorts of argument, before we relate our results to other theories of reasoning.

One criticism of our studies is that the materials are artificial, contrived, or ambiguous. In our view, the criticism has little merit. Consider a typical conditional used in our experiments:

> If Andrés played a game then he didn't play soccer.

In comparison with assertions, such as 'If there's an "A" on one side of card then there's a "2" on the other side of the card', the present example seems neither contrived nor artificial. Likewise, it seems no more ambiguous than many of the conditionals that occur in daily life. The crux of our rebuttal, however, is that the difference between the contents of the paradoxes and of the modulated inferences—at least with respect to artificiality, contrivedness, and ambiguity—is minimal. The two sorts of conditional typically differed by just a single word, e.g.:

> If Andrés played a game then he didn't play music.

Yet, the difference yielded a large and robust difference in the acceptance of the two sorts of conclusion. A skeptic might argue that individuals are likely to accept at least some inferences in such experiments, and we accept this claim. What it fails to explain is why the inferences that the participants accepted tended to be those with modulated conditionals. In summary, we see no obvious way to explain away our results in terms of flaws in the materials, or the design of the experiments.

A more general methodological argument against our studies proceeds as follows. The experiments show that naïve individuals reject the paradoxes, but accept the inferences with modulated conditionals. But, the latter have a different logical form from the former, and so the results have nothing to say about the paradoxes themselves. This argument, however, misses the point of our research strategy. We argue that the paradoxes are valid, but people reject them because their conclusions refer to possibilities inconsistent with the premises. Hence, if inferences have the same grammatical form, but contents that block the possibility inconsistent with the premise, then individuals should be more likely to accept the inferences. Our experiments corroborated this prediction. As we mentioned earlier, the strategy is akin to a study of the Müller-Lyer illusion, which illuminates its cause by comparing it to a configuration that no longer yields the illusion. To argue that the results of such a study tell us nothing about the Müller-Lyer would be wrong. They could corroborate a theory of the illusion. Likewise, we argue, our experiments corroborate the model theory of the paradoxes.

An additional point is pertinent. The model theory makes no use of logical form; and no algorithm exists for computing it for everyday conditionals. So, what *is* the logical form of a modulated conditional, such as, *if Lucia wore jewelry then she didn't wear the bracelet*. This conditional has the grammatical form, *if A then not B*, and it refers to two possibilities:

A	not-B
not-A	not-B

We can recover its logical form from these possibilities (and perhaps in no other way):

A or not A, and not B.

And so we can paraphrase the conditional as: Lucia either wore jewelry or not, but in either case she didn't wear the bracelet. If an inference is valid when it refers, say, to two possibilities, then it remains valid with the addition of any further possibility. Hence, in the present case, it remains valid when the conditional conclusion refers to these three possibilities:

A	not-B
not-A	not-B
not-A	B

But, they correspond to the possibilities to which a paradoxical conditional refers, and so that inference is valid too. Validity seems transparent in the case of the disjunctive paradoxes; and the preceding argument makes the case for the validity of the conditional paradoxes. When modulation blocks the construction of possibilities that conflict with the premises, individuals accept the inferences. The model theory predicted the results, but they present a challenge to three main sorts of theory: theories based on suppositions, theories based on probabilities, and theories based on formal rules of inference.

Suppositional theories derive from a footnote in Ramsey (1929/1990, p. 155) to the effect that the way that individuals establish their degrees of belief in a conditional is to add the contents of its if-clause hypothetically to their beliefs, and then to argue on that basis about the proposition expressed in the then-clause. Ramsey was writing, not about the meaning of a conditional, which he took to be material implication, but about how individuals fix their degrees of belief in a conditional. Stalnaker (1968), however, took Ramsey's account, and generalized it in order to frame a theory of the meaning of conditionals; and more recently Evans and Over (2004) have defended a version of this account as a psychological theory. Suppositional accounts reject the validity of the paradoxes (see our account in the Introduction). Evans and Over (2004, p. 19) wrote: 'The "paradoxes" are the absurd consequences of claiming that *ordinary* conditionals are truth functional' (p. 19, their italics). They also commented: 'We hold that the paradoxes are logically invalid for all natural language conditionals' (p. 153). Given their absurdity and invalidity on this account, it is no surprise that individuals tend to reject the paradoxes. But, why do they accept them with modulated conditionals, which eliminate possibilities inconsistent with the premises? If the paradoxes are invalid for *all* natural language conditionals, as the quotation above asserts, the suppositional account has no answer to this question. In our view, the suppositional account is plausible, and it is compatible with the model theory, which posits that individuals can base inferences on suppositions (Johnson-Laird and Byrne, 2002, p. 667). The difference the between the two theories on the paradoxes concerns merely an additional and unnecessary assumption of the suppositional theory—that ordinary conditionals are never truth-functional. The burden of our results is that some conditionals *are* truth functional, just as some conjunctions and disjunctions are.

Probabilistic theories give various accounts of validity (see the contrast between Oaksford and Chater, 2007 and Pfeifer and Kleiter, 2005). And at least one proponent of such a theory, Adams (1975), argues for a probabilistic conception of validity according to which the paradoxes are invalid. His theory does not offer any immediate account—let alone a prediction—of our results (see also Edgington, 1995). However, feasible probabilistic accounts of our results are due to Klaus Oberauer (personal communication, September, 2010) and to Mike Oaksford and Nick Chater (personal communications, July 2010, August 2011). Oberauer argues that individuals tend to accept a conclusion, such as:

If it rains today then the forecast is wrong,

if the conditional probability, p(forecast wrong | rain today) is high. They know that 'the forecast' refers to the forecast that it won't rain today (as a result of taking context into account), and a conditional is accepted to the degree that the conditional probability of the consequent given the antecedent is high. The conclusion gains such a probability only by virtue of the premise that disambiguates the expression 'this forecast'. Granted this disambiguation, the conditional is true. And that suffices to explain its acceptability. This post hoc explanation runs in parallel to the model theory, including the use of context to yield the inference that the 'this forecast' refers to the forecast that it won't rain. Pfeifer and Kleiter (2011) report an experiment investigating the two sorts of conditional paradox: the majority of the participants identified both sorts of paradox as 'probabilistically non-informative', i.e. in accordance with their probabilistic semantics, but not in accordance with material implication. These results are consistent with our theory, because these authors did not use semantic modulation to eliminate the possibility in the conditional conclusion inconsistent with the premises.

One apparent difficulty for probabilistic accounts occurs in the case of the disjunctive paradoxes. Given *B*, the probability of *A or B*, should normally approach unity (as Oberauer concedes). Oaksford, however, offers the following explanation of the effects of modulation (personal communication, August 2011). A disjunctive conclusion will be acceptable only if the first disjunct does not decrease the probability of the second disjunct. In a paradoxical inference, such as:

Eva read a newspaper.
Therefore, Eva read Don Quixote or she read a newspaper,

the truth of the first disjunct reduces the probability of the second disjunct to zero, and so it is unacceptable. In the modulated inference, such as:

> Eva read a novel.
> Therefore, Eva read Don Quixote or she read a novel,

the truth of the first disjunct increases the probability of the second disjunct, and so it is acceptable. The same account applies mutatis mutandis to the conditional paradox. It also runs in parallel to the model theory. The model theory argues that the disjunctive paradoxes are acceptable if a true first clause implies a true second clause; the probabilistic theory argues that they are acceptable if a true first clause raises the probability that the second clause is true. So, can we distinguish between the two accounts? The answer will depend on whether a mere rise in the probability of a proposition, as opposed to establishing its truth, suffices to make a paradoxical inference acceptable.

Formal rule theories allow that the paradoxes are valid, and Rips (1994, p. 47, p. 156) describes various putative explanations of their unacceptability, including Grice's (e.g. 1989) well-known pragmatic argument that a speaker who knows a categorical proposition that implies a conditional or a disjunctive conclusion is uncooperative to assert them instead of the categorical. On this account, the paradoxes are odd because their premises are more informative than their conclusions, which throw information away. Formal rules theories can therefore explain the unacceptability of the paradoxes. However, they have no machinery for reference to possibilities or for modulation, and so they cannot explain the acceptability of the modulated inferences. These theories also postulate that inference depends on the extraction of the logical form of premises, but how this process is carried out is itself a profound mystery, because—as our modulated assertions show—the grammatical form of a sentence can be remote from its logical form. Indeed, such is the size of this gap that some logicians are skeptical about the relevance of logical form to natural language (Barwise, 1989, p. 4).

We conclude that human reasoners diverge from standard logic in two crucial ways: they reject valid conclusions that throw information away by adding possibilities inconsistent with the premises—just as they reject other sorts of valid inference (see, e.g., Johnson-Laird, 2006, p. 157), and they use their knowledge to modulate the core meanings of connectives so that they are no longer interpreted in a strictly logical, truth-functional way. The logical interpretation is equivalent to material implication (or material equivalence) and yields the valid but unacceptable paradoxes. The modulated interpretation yields valid and acceptable inferences. The model theory predicts these results, but we make no claim that it is unique in this respect.

Department of Psychology, Universidad de La Laguna

Department of Psychology, University of Princeton

Appendix A: Materials for Experiment 1 (translated from the Spanish)

Modulated contents for the *Not B. therefore if A then not B* inferences.

1. Andrés didn't play soccer. Does it follow that if Andrés played a game then he didn't played soccer?
2. Javier didn't eat the apple. Does it follow that if Javier ate the fruit then he didn't eat the apple?
3. Maria didn't get the table. Does it follow that if Maria got the furniture then she didn't get the table?
4. Lucia didn't wear the bracelet. Does it follow that if Lucia wore jewelry then she didn't wear the bracelet?
5. Paco didn't buy the chicken. Does it follow that if Paco bought meat then he didn't buy the chicken?

Paradoxical contents for the *Not B. therefore if A then not B* inferences.

6. Luisa didn't play music. Does it follow that if Luisa played a game then she didn't play music?
7. Veronica didn't eat the lentils. Does it follow that if Veronica ate the fruit then she didn't eat the lentils?
8. José didn't get the umbrella. Does it follow that if José got the furniture then he didn't get the umbrella?
9. Marcos didn't wear the shoes. Does it follow that if Marcos wore jewelry then he didn't wear the shoes?
10. Encarna didn't buy the sardines. Does it follow that if Encarna bought meat then she didn't buy the sardines?

Modulated contents for the *Not A. therefore if A then B* inferences.

11. It won't rain today. Does it follow that if it does rain today then the forecast is wrong?
12. The experiment won't work. Does it follow that if the experiment works then this prediction is false?
13. The politicians won't pass the budget. Does it follow that if the politicians do pass the budget then many of them changed their views?
14. Lorraine will win her tennis match. Does it follow that if Lorraine doesn't win her tennis match then this expectation is incorrect?
15. Lawrence is going on strike. Does it follow that if Lawrence doesn't go on strike then his colleagues were mistaken about him?

Paradoxical contents for the *Not A. therefore if A then B* inferences.

16. It won't rain today. Does it follow that if it does rain today then the forecast is right?
17. The experiment won't work. Does it follow that if the experiment does work then the prediction is true?
18. The politicians won't pass the budget. Does it follow that if the politicians do pass the budget then many of them stayed constant to their views?
19. Lorraine will win her tennis match. Does it follow that if Lorraine doesn't win her tennis match then this expectation was correct?
20. Lawrence is going on strike. Does it follow that if Lawrence doesn't go on strike then his colleagues were right about him?

Appendix B: Materials for Experiment 2 (translated from the Spanish)

Modulated contents for disjunctive inferences

1. Paco visited France. Does it follow that Paco visited Paris or he visited France?
2. Ana read a novel. Does it follow that Ana read Don Quixote or she read a novel?
3. Pedro tried the dessert. Does it follow that Pedro tried ‘the chocolate cake’ or he tried the dessert?
4. Lucia watched the football. Does it follow that Lucia watched the Barca or she watched the football?
5. Andres drank the wine. Does it follow that Andrés drank wine ‘Jumilla’ or he drank wine?

Paradoxical contents for disjunctive inferences

6. David visited England. Does it follow that David visited Paris or he visited England?
7. Eva read a newspaper. Does it follow that Eva read Don Quixote or she read a newspaper?
8. Gorka tried the jam. Does it follow that Gorka tried ‘the chocolate cake’ or he tried the jam?
9. Alba watched the tennis. Does it follow that Alba watched the Barca or she watched the tennis?
10. Carlos drank a coke. Does it follow that Carlos drank wine ‘Jumilla’ or he drank a coke?

Modulated contents for conditional inferences

11. Rocio didn’t buy oil ‘Carbonell’. Does it follow that if Rocio bought oil then she didn’t buy oil ‘Carbonell’?
12. Maria didn’t listen to pop music. Does it follow that if Maria listened to music then she didn’t listen to pop music?
13. Julio didn’t watch *The Matrix*. Does it follow that if Julio watched a movie then he didn’t watch *The Matrix*?
14. Fanny didn’t travel by tram. Does it follow that if Fanny travelled by public transport then she didn’t travel by tram?
15. Pepe didn’t buy a Mercedes. Does it follow that if Pepe bought a car then he didn’t buy a Mercedes?

Paradoxical contents for conditional inferences

16. Jeni didn’t buy sugar. Does it follow that if Jeni bought oil then she didn’t buy sugar?
17. Olga didn’t listen to a talk. Does it follow that if Olga listened to music then she didn’t listen to a talk?
18. Lucas didn’t watch the news. Does it follow that if Lucas watched a movie then he didn’t watch the news?
19. Isa didn’t travel by foot. Does it follow that if Isa travelled by public transport then she didn’t travel by foot?
20. Juan didn’t buy a house. Does it follow that if Juan bought a car then he didn’t buy a house?

References

Adams, E. W. 1975: *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Dordrecht: Reidel.

Bar-Hillel, Y. and Carnap, R. 1964: An outline of a theory of semantic information. In Y. Bar-Hillel (ed.), *Language and Information*. Reading, MA: Addison-Wesley.

Barres, P. and Johnson-Laird, P. N. 2003: On imagining what is true (and what is false). *Thinking & Reasoning*, 9, 1–42.

Barrouillet, P., Grosset, N. and Lecas, J. F. 2000: Conditional reasoning by mental models: chronometric and developmental evidence. *Cognition*, 75, 237–66.

Barwise, J. 1989: *The Situation in Logic*. Stanford, CA: Center for the Study of language and Information.

Bonnefon, J. F. and Politzer, G. 2011: Pragmatic, mental models, and one paradox of the material conditional. *Mind & Language*, 26, 141–55.

Byrne, R. M. J. and Johnson-Laird, P. N. 2009: 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences*, 13, 282–6.

Edgington, D. 1995: On conditionals. *Mind*, 104, 235–329.

Evans, J. St. B. T. and Over, D. E. 2004: *If*. Oxford: Oxford University Press.

Gregory, R. L. 1997: *Eye and Brain*, 5th edn. Princeton, NJ: Princeton University Press.

Grice, H. P. 1989: *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Jeffrey, R. 1981: *Formal Logic: Its Scope and Limits*. New York: McGraw-Hill.

Johnson-Laird, P. N. 1983: *Mental Models*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. 2006: *How We Reason*. Oxford: Oxford University Press.

Johnson-Laird, P. N. and Byrne, R. M. J. 1991: *Deduction*. Hove: Psychology Press.

Johnson-Laird, P. N. and Byrne, R. M. J. 2002: Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646–78.

Oaksford, M. and Chater, N. 2007: *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

Oberauer, K. 2006: Reasoning with conditionals: a test of formal models of four theories. *Cognitive Psychology*, 53, 238–83.

Peirce, C. S. 1931–1958: Volume 4. In C. Hartshorne, P. Weiss and A. Burks (eds), *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Pfeifer, N. and Kleiter, G. D. 2005: Towards a mental probability logic. *Psychologica Belgica*, 45, 71–99.

Pfeifer, N. and Kleiter, G. D. 2011: Uncertain deductive reasoning. In K. I. Manktelow, D. E. Over and S. Elqayam (eds), *The Science of Reason: A Festschrift in Honour of Jonathan St. B. T. Evans*. Hove: Psychology Press, 145–66.

Quelhas, A. C., Juhos, C. and Johnson-Laird, P. N. 2010: The modulation of conditional assertions and its effects on reasoning. *Quarterly Journal of Experimental Psychology*, 63, 1716–39.

Ramsey, F. P. [1929] 1990: General propositions and causality. In D. H. Mellor (ed.), *F. P Ramsey: Philosophical Papers*. Cambridge: Cambridge University Press, 145–63.

Rips, L. J. 1994: *The Psychology of Proof*. Cambridge, MA: MIT Press.

Rips, L. J. and Conrad, F. G. 1983: Individual differences in deduction. *Cognition and Brain Theory*, 6, 259–285.

Santamaría, C., García-Madruga, J. A. and Johnson-Laird, P. N. 1998: Reasoning from double conditionals: the effects of logical structure and believability. *Thinking & Reasoning*, 4, 97–122.

Schroyens, W. 2009: Mistaking the instance for the rule: a critical analysis of the truth-table evaluation paradigm. *Quarterly Journal of Experimental Psychology*, 63, 246–59.

Schroyens, W. 2010: Logic and/in psychology: the paradoxes of material implication and psychologism in the cognitive science of human reasoning. In M. Oaksford and N. Chater (eds), *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford: Oxford University Press, 69–84.

Sperber, D. and Wilson, D. 1995: *Relevance: Communication and Cognition*, 2ndedn. Oxford: Basil Blackwell.

Stalnaker, R. C. 1968: A theory of conditionals. In N. Rescher (ed.), *Studies in Logical Theory*, American Philosophical Quarterly Monograph No. 2. Oxford: Blackwell, 98–122.

Stalnaker, R. C. 1975: Indicative conditionals. *Philosophia*, 5, 269–86.

Verschueren, N., Schaeken, W. and d'Ydewalle, G. 2005: A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, 11, 278–93.