HOW IMPLICATION IS UNDERSTOOD

By P. N. JOHNSON-LAIRD and JOANNA TAGART,
University College London, England

Conditional sentences of the form "if p then q" are often difficult to evaluate, as students of logic well know. The conditional is clearly true when both antecedent (p) and consequent (q) are true, and false when the antecedent is true but the consequent false. But what truth-value should be assigned when the antecedent is false? Logicians, working with a propositional calculus that permits only values of truth or falsity, stipulate that the implication is true in this case—regardless of the truth-value of the consequent. However, it seems that the conditional sentence might fail to do justice to this notion of *material* implication.

Conditional sentences have been found to present difficulties to both children and adults. Matalon and Peel suggest that children tend to interpret the conditional as a material equivalence ("if and only if p then q"), which is true when antecedent and consequent have the same truth-value and false when their values are different.[1] Wason, however, argues that adults do not treat the conditional in a truth-functional manner: they consider it to be *irrelevant* when its antecedent is false.[2] For example, when someone says, "if it's raining, then I'm going to the cinema," the statement is neither true nor false but merely irrelevant if in fact it is *not* raining. The layman is unlikely to consider the statement to be true just because it is not raining. Indeed, logicians have long recognized, and argued about, this way of interpreting conditional sentences.[3]

The question naturally arises as to the extent to which the inter-

* Received for publication August 20, 1968. The authors express their gratitude to Dr. P. C. Wason for his advice and encouragement, and for a critical reading of an earlier version of this paper.

[1] B. Matalon, Etude genetique de l'implication, *Etudes d'epistemologie genetique: XVI, Implication, formalisation et logique naturelle,* 1962, 69–95; E. A. Peel, A method for investigating children's understanding of certain logical connectives used in binary propositional thinking, *Brit. J. math. statist. Psychol.,* 20, 1967, 81–92.

[2] P. C. Wason, Reasoning, in B. M. Foss (ed.), *New Horizons in Psychology,* 1966, 135–151.

[3] W. V. O. Quine, *Methods of Logic,* rev. ed., 1956, 12; W. Kneale and M. Kneale, *The Development of Logic,* 1962, 128–138.

pretation of implication is influenced by the manner of its expression. The present experiment was designed to examine the effect of statements of these types:

1. If p then q.
2. There isn't p, if there isn't q.
3. Either there isn't p, or there is q (or both).
4. There is never p without there being q.

To the logician, these sentences may be interpreted as expressing material implication.[4] On Wason's hypothesis, however, we expected that Sentence 1 would be considered irrelevant to any situation that falsified its antecedent. Similarly, Sentence 2, which was derived from the contrapositive "if not q, then not p," would also be considered irrelevant when its antecedent is false, *i.e.* when q is true. But Sentences 3 and 4 were not conditionals: they lack the conditional term *if*, and it was predicted that they would be treated as expressing material implication and elicit fewer judgments of irrelevancy. It will be noted that Sentence 3 contains the term *either* which tends to suggest exclusive disjunction[5] but that this is countermanded by the presence of *or both*, which is an explicit statement of inclusive disjunction. The predicted classifications are summarized in Table I.

There are a number of ways in which an antecedent or consequent may be falsified. For example, an antecedent like "if there's a letter A," which was used in the present experiment, can be falsified by the occurrence of a letter B, or of a geometrical shape, or of nothing whatsoever. Logically, these would be equivalent falsifications; psychologically, they might not be equivalent. This point was examined by using a suitable selection of stimuli.

TABLE I

PREDICTED CLASSIFICATIONS OF THE FOUR TYPES OF SENTENCE

	Situation			
Sentence type	pq	$p\bar{q}$	$\bar{p}q$	$\bar{p}\bar{q}$
1. If p then q.	T	F	?	?
2. Not-p if not-q.	?	F	?	T
3. Not-p or q.	T	F	T	T
4. Never p without q.	T	F	T	T

Note: T, F, and ?, respectively, denote judgments of truth, falsity, and irrelevance; $\bar{p}$ denotes not-p.

[4] P. F. Strawson, *Introduction to Logical Theory*, 1952, 35–40.

[5] A. Naess, L'emploi de la disjonction chez les adolescents, *Etudes d'epistemologie genetique: XVI, Implication, formalisation et logique naturelle*, 1962, 151–158.

METHOD

Each *S* was shown an array of stimuli and a sentence referring to them. The *S*'s task was to consider each stimulus in turn and to decide whether it indicated that the sentence was true or false, or was irrelevant to the truth-value of the sentence. *S*s served as their own controls and performed the task with the four different sentences expressing implication. The order of presentation of the sentences was counterbalanced over *S*s: each of the 4! possible orders was used once.

Materials. The sentences presented were variations of the four basic types: *e.g.* (1) If there is an A on the left, then there is a 7 on the right. (2) There isn't an A on the left, if there isn't a 7 on the right. (3) Either there isn't an A on the left, or there is a 7 on the right (or both). (4) There is never an A on the left without there being a 7 on the right. In order to reduce residual effects, the numbers and letters in each sentence that *S* received, and the corresponding arrays of stimuli, were different. There were four such sets of material, involving different letters and numbers, and their order of presentation was independently counterbalanced over the *S*s. The 16 sentences (4 types $\times$ 4 contents) were typed in capital letters on separate 6 $\times$ 2 in. cards.

Each set of stimuli consisted of 16 4 $\times$ 2 in. cards which were divided into two halves by a heavy ink line. On the left of the line there was a letter (either the one mentioned in the sentence or one other letter), or a geometrical shape, or nothing whatsoever; on the right of the line there was a number (either the one mentioned in the sentence or one other number), or the geometrical shape, or nothing whatsoever. Four such sets of stimuli were constructed with different numbers, letters, and shapes.

Procedure. The first set of stimuli was spread out in an arbitrary array in front of *S*, and the general purpose of the experiment was described. The *S* was told that although any stimulus must fall into one of the three classificatory categories ('true,' 'false,' 'irrelevant'), he must not assume that there would necessarily be cards in all three categories. When *S* understood what he had to do, the stimuli were gathered together and shuffled. *S* was told that he was going to classify the cards one at a time, each card being placed in an appropriate pile, and that he would be timed. *S* was timed from the moment that he received the sentence until he had completed the classification of the stimuli. The sentence remained on view throughout the classification. The subsequent classifications followed the same procedure except that the initial display of stimuli was omitted.

Subjects. Twenty-four *S*s were individually tested. They were all students at University College London, and native speakers of English.

RESULTS

There were four stimuli which were crucial in evaluating the results. They consisted, for a sentence of the form "if there's an A . . ., then there's a 7 . . .," of the items A7, A8, B7, and B8, *i.e.* pq, $p\bar{q}$, $\bar{p}q$, and $\bar{p}\bar{q}$. Table II shows the six common classifications of these stimuli, and the frequencies with which they occurred for each type of sentence. None of the remaining classifications occurred

more than twice throughout the whole experiment. There were 28 different types of classification altogether, 6 common and 22 miscellaneous ones, out of a total of 81 possible classifications.

The most frequent classification, for Sentences 1, 2, and 3, was the predicted one; and the actual frequencies were all significant ($p < .01$), assuming independent classifications, on binomial tests based conservatively upon the actual number of different types of classification for each sentence. Contrary to expectation, Sentence 4 tended to be classified in the same way as Sentence 1. Fifteen *S*s produced the same classification for these two sentences, whereas there were only four other occasions when an *S* produced the same classification for two sentences. The chance probability of obtaining the same classification for Sentences 1 and 4, with 2 and 3 being different and different from one another, is conservatively 1/28. Clearly, the similarity of the classifications of Sentences 1 and 4 was not due to chance.

The mean number of 'irrelevant' judgments of the 16 stimuli and the mean classification times are given in Table III. The difference between the sentences of the number of 'irrelevant' judgments was significant on a Friedman analysis of variance ($\chi^2_r = 32.3$, $p < .001$). Sentences 1 and 4 tended to elicit 'irrelevant' judgments when their antecedents were false, and this was also the case to a lesser extent for Sentence 2. There was no tendency in any condition for 'irrelevant' judgments to increase when falsification was due to a geometrical shape or a 'blank' rather than to a letter or a number. On this point the *S*s' rationality was vindicated.

The classification times for the four sentences were also significantly different on a Friedman analysis of variance ($\chi^2_r = 34.5$,

TABLE II

CLASSIFICATIONS OF THE FOUR CRUCIAL STIMULI AND CLASSIFICATION FREQUENCIES OF OCCURRENCE FOR EACH TYPE OF SENTENCE

	Stimuli				Sentence type				
	pq	$p\bar{q}$	$\bar{p}q$	$\bar{p}\bar{q}$	1	2	3	4	Total
Common classifications	T	F	?	?	19	1		14	34
	T	F	T	T	1	1	8	3	13
	?	F	?	T		5			5
	F	F	T	T		2	4		6
	F	F	T	?		1	2		3
	T	F	?	T		2		1	3
Miscellaneous classifications					4	12	10	6	32
Totals					24	24	24	24	96

Note: Miscellaneous classifications are those which did not occur more than twice throughout the whole experiment.

TABLE III

MEAN NUMBER OF 'IRRELEVANT' JUDGMENTS AND MEAN CLASSIFICATION TIMES FOR EACH TYPE OF SENTENCE

	Sentence type			
	1	2	3	4
'Irrelevant' judgments	10.2	5.1	1.6	9.3
Classification times (in sec.)	45	96	107	60

$p < .001$). Although there was a significant learning effect for classification times ($p < .003$, Jonckheere group test for predicted trend), there was no such effect for logical accuracy.

DISCUSSION

This experiment showed that the way in which implication is expressed exerts a decisive influence upon what it is understood to denote. When expressed in the form of a conditional "if p then q" or "not p if not q" it was, as predicted, treated in a non-truth-functional manner. Unexpectedly, the same interpretation—in which stimuli falsifying the antecedent were regarded as irrelevant—was elicted by the sentence "there is never p without q." Hence, the term *if* is by no means necessary to elicit the non-truth-functional interpretation; and in the absence of an account of the semantics of these sentences, such necessary conditions remain obscure. Similarily, it seems likely that *if* cannot be taken as an unequivocal marker of the antecedent of conditionals: a sentence of the form "p only if q" is likely to receive the same interpretation as "if p then q." To what extent is the non-truth-functional interpretation due to the implicit invitation to classify stimuli as irrelevant? Performance on the disjunctive sentence suggests that *S*s were able to resist the 'irrelevant' category, and it is plausible to assume that the classifications did reflect the spontaneous interpretations of the sentences. Likewise, it seems improbable that the specific content of the sentences and stimuli should have exerted any major distorting influence upon performance.

The sentence which was most often classified as material implication was the disjunction. This was never treated as exclusive disjunction, but an interesting error proved most persistent. Such was the force of the phrase "either there isn't p . . ." that a number of Ss produced a truth-functional classification appropriate to the simple proposition "not p" (see Table II).

The manner in which the conditionals were interpreted, considered

in conjunction with the findings of Matalon and Peel,[6] raises certain difficulties for Piaget's account of intellectual development. Preadolescent children tend to treat conditionals as expressions of material equivalence; adolescents at the level of propositional operations treat them as expressions of material implication; yet, undergraduates in the present experiment failed to treat them as any sort of truth-functional connective. Piaget believes that an individual tests a putative causal relation by expressing it in the form "if p then q" and then searching for its counterexample, formed by negating the material implication.[7] But adults evidently do not readily interpret "if p then q" as material implication. Even if they did, further doubt is cast upon Piaget's position by an unpublished experiment by Wason and Johnson-Laird, in which subjects were presented with four cards bearing values of p, $\bar{p}$, q, and $\bar{q}$. They were told that every card had a value of p or $\bar{p}$ on one side, and q or $\bar{q}$ on the other side; and they were asked to choose those cards which it was necessary to turn over to test whether a given conditional rule was true or false. There was, indeed, a consistent tendency for subjects to choose the cards which fulfilled the antecedent: p in the case of "if p then q," and $\bar{q}$ in the case of "if not q then not p." However, subjects were reluctant to choose those cards which falsified the consequent, especially in the case of "if p then q." Yet such cards are a required choice on any reasonable interpretation of the conditional, including even the non-truth-functional interpretation of the present experiment.

Such a result makes a stark contrast with Piaget's views, and with the findings of Stewart and Hill that adults and children correctly evaluate inferences of the form "if p then q; not q, therefore not p."[8] It would seem therefore that there are crucial psychological differences between *making* inferences and merely *evaluating* them. Not only do *S*s fail to make the inference in the card-turning test, but their failure, as Wason has shown, is resistant to a number of "therapeutic" procedures.[9]

Finally, we may ask how implication is best expressed in the

[6] Matalon, *loc. cit.*; Peel, *loc. cit.*

[7] E. W. Beth and J. Piaget, *Mathematical Epistemology and Psychology*, 1966, 181.

[8] D. K. Stewart, Communication and logic: Evidence for the existence of validity patterns, *J. gen. Psychol.*, 64, 1961, 297–305; S. Hill's findings reported by P. Suppes, On the behavioral foundations of mathematical concepts, in L. N. Morrisett and J. Vinsonhaler (eds.), Mathematical learning, *Monogr. Soc. Res. Child Dev.*, 30, 1965 (No. 1), 60–96.

[9] P. C. Wason, Reasoning about a rule, *Quart. J. exp. Psychol.*, 20, 1968, 273–281.

English language. There is no readily available answer to this question: we are faced with a dilemma. On the one hand, disjunction yields an implicational interpretation more often than the conditional sentences, but it takes longer to process and has a tendency to produce diverse and labile interpretations—a finding which has been recently confirmed.[10] And such ambiguities are likely to be reflected in tasks involving the evaluation of inferences.[11] On the other hand, performance with "if p then q," though faster and more stable, is not consistent with material implication. However, this departure from the logicians' calculus has an unexpected advantage. It breaks the logical relation between the conditional and its contrapositive: they no longer imply one another. This does away with the paradoxes of material implication[12] and the paradoxes of confirmation,[13] at least for the conditionals of everyday language.

Summary

Students classified stimuli according to whether they indicated that a sentence was true or false, or were irrelevant to the truth-value of the sentence. Four different sentences were used, with *Ss* acting as their own controls, and each sentence was logically equivalent to a material implication. The results showed that disjunction ("not-p or q") yielded the greatest number of classifications in accordance with the truth-values of implication. The remaining sentences ("if p then q," "not-p if not-q," "never p without q") were not classified in a truth-functional way: stimuli were judged irrelevant when they falsified the antecedents of these sentences. The results would seem to raise some difficulties for Piaget's notion of the developmental level of formal operations.

[10] P. C. Wason and P. N. Johnson-Laird, Proving a disjunctive rule, *Quart. J. exp. psychol.*, 21, 1969, 14–20.

[11] P. N. Johnson-Laird, On understanding logically complex sentences, *Quart. J. exp. Psychol.*, 21, 1969, 1–13; P. N. Johnson-Laird, Reasoning with ambiguous sentences, *Brit. J. Psychol.*, 60, 1969, 17–23.

[12] Strawson, *op. cit.*, 88.

[13] C. I. Hempel, Studies in the logic of confirmation, *Mind*, 54, 1945, 1–26, 97–121, reprinted in C. I. Hempel, *Aspects of Scientific Explanation*, 1965, 3–46.