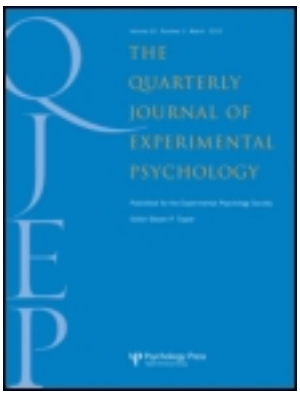

Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje19

Insight into a logical relation

P. N. Johnson-Laird [a] & P. C. Wason [a]

[a] Department of Psychology, Psycholinguistics Research Unit, University College, London

Version of record first published: 29 May 2007.

To cite this article: P. N. Johnson-Laird & P. C. Wason (1970): Insight into a logical relation, Quarterly Journal of Experimental Psychology, 22:1, 49-61

To link to this article: http://dx.doi.org/10.1080/14640747008401901

Q. Jl exp. Psychol. (1970) **22**, 49–61

INSIGHT INTO A LOGICAL RELATION

P. N. JOHNSON-LAIRD AND P. C. WASON

Department of Psychology, and Psycholinguistics Research Unit, University College London

Two experiments are reported which aimed to investigate factors affecting the gain of insight into the logical relation of implication. In the first experiment, subjects had to make a series of inferences about either a conditional sentence or a quantified sentence, both of which had the same underlying logical form. Under one condition the sentences had to be proved true, and under another condition, false. Proving a sentence false facilitated gain of insight, but the linguistic form of the sentence exerted no significant effect on the main dependent variable. In the second experiment, implication was not expressed as a sentence but was inherent in the structure of the task. The experimental material differed in complexity and allowed the cognitive load imposed on the subject to be varied. Results suggested that insight was not all-or-none. It was spontaneously gained when the material was simple, but temporarily lost when it was complex.

Introduction

Previous experiments (Wason, 1968) have shown that in an abstract task people reason inadequately about conditional sentences. For example, imagine this experiment. The subject is given the sentence, "if there is a D on one side of a card, then there is a 3 on the other side," together with four cards which show face upwards D, K, 3 and 7, respectively. He knows that there is always a letter on one side of a card and a number on the other side. The task is to say which of the cards need to be turned over in order to find out whether the sentence is true or false. The correct answer is, "the D and the 7," but hardly any subjects will say that the 7 needs to be turned. And yet it does need to be turned because a D on its other side would falsify the sentence in just the same way as would a 7 (or any digit other than 3) on the other side of the D. This error of omission is the result of a deep fixation; although a more recent experiment (Wason, 1969) showed that confronting the subjects with a series of contradictions does facilitate a subsequent correction.

The present investigation was designed to answer two questions: (1) whether insight into the relation of implication would be more readily gained if the task were both concrete and simplified, and (2) whether such insight is an all-or-none matter, or whether it would vary over a series of trials as a function of the cognitive load imposed.

Experiment I

This experiment was designed primarily to examine the grasp of implication in a simplified, concrete task. But it also aimed to investigate the effects of varying (a) a syntactic factor and (b) a semantic factor.

The syntactic factor concerned the difference between a sentence of the form, "if they are p's then they are q" (conditional) and one in the form, "all p's are q" (quantified). Sentences in these forms differ linguistically but have the same underyling logical structure. H. H. Clark (personal communication) has pointed out that conditional sentences in everyday life are frequently construed as causal or temporal, and that the previous results (Wason, 1968) might have reflected this factor.

The semantic factor consisted in the difference between proving a sentence true and proving it false. When implication is expressed as negated disjunction, "either not-p or q," this factor has no effect (Wason and Johnson-Laird, 1969), but when it is expressed in conditional or quantified form it would seem that proving the sentence false might well facilitate the correct response of selecting "not-q."

The subjects were presented with a sentence that purported to describe the contents of two boxes: one labelled "white" and the other labelled "black." The sentence was either, "*all the triangles are black*" (quantified), or "*if they are triangles, then they are black*" (conditional). The task was to attempt to prove, as efficiently as possible, in one condition that the sentence was true, and in another condition that it was false. There were thus four different tasks. The establishment of proof was to be done by examining the contents of the boxes. At every trial the subject requested a stimulus from either the white box, or the black box, but he did not know what its shape would be except that the contents of the boxes were limited to triangles and circles. The correct response, whether the sentence is to be proved true or false, is to choose only white stimuli. It is only the presence of a white triangle which could prove the sentence false, and it is only the complete absence of one which could prove it true. Hence the contents of the black box are entirely superfluous, but the contents of the white box must be exhaustively examined. However, since intuitively it seems more difficult to appreciate that a sentence must be proved true by establishing the non-existence of a stimulus, it was predicted that more black stimuli would be requested when the sentence had to be proved true than when it had to be proved false. Similarly, it was predicted that more black stimuli would be requested when the sentence was conditional in form than when it was quantified. Logically neither sentence asserts the existence of black triangles, but the conditional was expected to arouse curiousity to see whether triangles exist. And the natural place to look for them is "obviously" in the black box.

Design

Four independent groups: a group who had to prove the conditional sentence true, and a group who had to prove it false; a group who had to prove the quantified sentence true, and a group who had to prove it false.

Subjects

Twenty-eight first-year psychology undergraduates of University College London were allocated in rotation to the groups and tested individually. There were thus seven subjects in each group.

Procedure

The subjects were told that they would carry out a reasoning task which was not a test of their intelligence. The appropriate sentence was placed on the desk, typed on a card (6 × 4 in.), together with two cardboard boxes (4 × 3 × 2½ in.), open at the back, and labelled "White" and "Black," on the subjects' left and right, respectively. The subjects were told that their task was to prove the sentence true (or false); and that all the objects in the white box were white and all those in the black box were black. They were also told that they could predict nothing about their shape except that only triangles and circles were used.

Their task was explained. "You ask me to hand you a shape from either box, and then, when you have examined it, you ask me to hand you another shape—again from either box. We continue in this way until you are satisfied you have proved the sentence true (false). There is no hurry, but I want you to do the task economically, i.e. to examine as few shapes as you need. There are exactly 15 shapes in each box."

The stimuli consisted of black and white triangles and circles of different sizes cut out of cardboard. The contents of the black box were the same for all four groups: 10 triangles and 5 circles. The contents of the white box were the same for the two groups who had to prove the sentence true: 15 circles; and the same for the two groups who had to prove it false: 14 circles and 1 triangle. In other words, the sentence was true for the groups who had to prove it true, and false for the groups who had to prove it false. But this would be apparent only if all the contents of the white box were examined.

The order in which the stimuli were presented was pre-determined. If a black shape were to be requested by the subject, it was assumed that initially presenting a circle, rather than a triangle, would immediately facilitate gain of insight into the task, because, in realizing that not all black stimuli were triangles, he might appreciate that the contents of the black box conveyed no information. The aim of the experiment was to see whether such insight would be *spontaneously* gained. Hence the presentation of a circle was delayed for five trials. Accordingly, 5 triangles were presented first of all in all four groups, followed by 1 circle and the remaining 8 stimuli in a random order. Insight, if it occurred, would have to overcome the "reinforcement" provided by the first 5 black triangles.

In the two groups who had to prove the sentence false the 14 white circles were presented first of all followed by the single white triangle. Thus insight into the task in these two groups was revealed by persistence in requesting white stimuli in spite of the negative "reinforcement" provided by 14 white circles.

When the stimuli had been presented they were not replaced. The subject was allowed to arrange them in front of him on the desk, and to count them. If at any point it was claimed that the sentence had been proved true, when it had not been proved, the subject was told he was wrong and asked to continue with the task.

Results

In all cases the contents of the white box were totally exhausted by the subjects, and in no cases were the contents of the black box totally exhausted. In other words, all the subjects evidently gained insight at some point. The most sensitive dependent variable is accordingly the total number of black shapes requested. Table I shows the individual scores and means for each of the four groups.

A Kruskal–Wallis one-way analysis of variance was computed on the data and was significant ($\chi^2 = 11{\cdot}36$, $d.f. = 3$, $P<0{\cdot}01$). The prediction that proving a sentence false, rather than true, would be easier, was confirmed. The overall mean for proving it true was 4·29 and the overall mean for proving it false was 0·86. The difference between the two conditions was highly significant ($S = 156$, $P = 0{\cdot}0002$, one-tailed). The prediction that the quantified sentence would

lead to more efficient performance than the conditional sentence was not confirmed, the overall means being respectively 2·07 and 3·07 ($S = 1$, N.S.).

Inspection of Table I shows that 21 out of the 28 subjects requested a white shape before encountering the first black circle, i.e. the sixth stimulus from the black box, and that only two subjects requested black stimuli for a further three trials.

TABLE I

Total number of black stimuli requested

	Prove true: quantified	Prove true: conditional	Prove false: quantified	Prove false: conditional
	3	6	2	0
	9	6	0	1
	2	9	0	0
	2	6	1	0
	1	1	2	6
	2	6	0	0
	5	2	0	0
Σ	24	36	5	7
$\bar{x}$	3·4	5·1	0·7	1·0

A more sensitive dependent variable, related to the temporal order in which the stimuli were requested, did, however, discriminate between the quantified and the conditional sentence. Disregarding the data from those subjects who showed perfect insight initially, i.e. those who never requested a black stimulus, the total number of white shapes requested *before* requesting any black shape was computed. The following distributions were obtained:

Quantified sentence: 0, 0, 0, 0, 1, 1, 1, 2, 4, 4 ($n = 10$, $\bar{x} = 1{\cdot}3$)

Conditional sentence: 0, 0, 0, 0, 0, 0, 0, 0, 1 ($n = 9$, $\bar{x} = 0{\cdot}1$)

The difference between these two conditions is significant ($S = 47$, $P = 0{\cdot}02$, one-tailed). These data seem to indicate a greater *readiness* on the part of those subjects, who had the conditional sentence, to request a black shape initially.

Discussion

Compared with the previous experiments (Wason, 1968), the present task differs in three important ways. It uses as material concrete stimuli, shapes in boxes, as opposed to symbols on both sides of a card. Second, inferences are made about only the consequent of the implication: the choice is between "white" and "black" rather than between "white," "black," "triangles" and "circles." Third, the sentence specifies a relation between the antecedent and consequent which coheres in a single stimulus, e.g. a "black triangle," as opposed to two conjoined stimuli, e.g. a D on one side of a card and a 3 on the other side.

These modifications did indeed make the task easier. When the instruction was to prove the sentence false, the task seemed almost trivial; only one subject

requested six black stimuli and none of the others requested more than two. And even when the instruction was to prove the sentence true, half the subjects requested no more than three black stimuli. Clearly, it was unnecessary to have as many as 15 stimuli in the boxes.

It is of interest to note that those subjects who had the quantified sentence did not request fewer black stimuli than those who had the conditional sentence. This suggests that it is not the difference between the linguistic form of these two sentences which is important but the underlying logical form which they share. This result has been corroborated by Johnson-Laird and Tagart (1969) who have shown that in an evaluation task there is no difference between sentences in the form, "if p then q" and "there is never p without there being q." Another experiment by Wason (1969) has shown that the quantified sentence, "every card which has a red triangle on one side has a blue circle on the other side," precludes initial insight even more frequently than do the conditional sentences used in the previous experiments. It is not linguistic form but logical form which is important.

This conclusion, however, must be qualified because there is a difference between the conditional and quantified sentence on the more sensitive dependent variable of the number of white stimuli requested before requesting any black stimuli. The quantified sentence evidently *delays* inspection of the black stimuli. This result may be related to the grounds on which the prediction, that the quantified sentence would lead to more efficient performance, is based. The conditional sentence may simply excite curiosity to establish the existence of triangles.

A deeper, and more speculative, hypothesis is connected with the corresponding negations of the two test sentences. The negation of the quantified sentence is "not all the triangles are black," but the negation of the conditional is not, of course, "if they are triangles, then they are not black," but "it is not the case that if they are triangles, then they are black." It is possible that the subjects may have covertly tried to formulate these propositions in order to guide their responses. The negation of the conditional is not only tortuous but seemingly vague, and it may have been the case that trying to find the words in which to formulate it resulted in such a confusing predicament that the subjects abandoned their efforts and fell back on that strong positive set, or direction, which is connoted by the unnegated conditional (Wason, 1968; Johnson-Laird and Tagart, 1969). Such a set would, of course, direct them towards the black box. The point, however, is a subtle one and the results show no gross differences in performance between the two kinds of sentence.

Experiment II

The task used in Experiment I is evidently too easy. What is wanted is a task in which the cognitive load imposed on the subject can be varied. In all the previous experiments the relation of implication has been expressed in a sentence about which the subject has to make inferences. It has been part of the experimental material.

In this experiment the relation of implication does not reside in the experimental material but is inherent in the structure of the task itself. To understand

the task is to understand the relation of implication. In contrast, the experimental material, which has to be understood, consists of sentences which vary in their logical complexity and thus allow cognitive load to be independently varied. The experiment aims to investigate the interaction between these two factors. It poses the question: is the appreciation of the task's implicative structure affected by the experimental material?

The subjects were presented with an array of stimuli (diagrams) which constituted the total universe of such stimuli as far as they were concerned. The nature of the problem that confronted them is best made clear by considering the instructions. The experimenter said: "I want you to imagine that I have taken some of these diagrams and put them in an envelope, sealed it, and then written a description of all the diagrams it contains. . . . Of course, I haven't put all the diagrams in the envelope and the description might also apply to some of the diagrams left outside the envelope. . . . Your task is to discover whether my description of the contents of the envelope is true or false. The way you will do this is by picking out, one at a time, those diagrams which you want information about. I will tell you whether each diagram you choose is inside or outside the envelope."

Consider a diagram which is a positive instance of a rule (i.e. a "description"). The instructions state that such a diagram may be either inside or outside the envelope. Hence knowledge of its location is vacuous. But consider a diagram which is a negative instance. If it is contained within the envelope, then the rule is decisively falsified; if it has been left outside the envelope, then the truth of the rule has been corroborated. Hence only negative instances are informative. The implicative structure of the task is now clear. From the sentence, "if it is inside the envelope, then it conforms to the rule," it does not follow that, "if it conforms to the rule, then it is inside the envelope," but what does follow is, "if it doesn't conform to the rule, then it isn't inside the envelope." And this has an obvious analogy to the material in Experiment I. In order to prove the sentence, "all the triangles are black," is true (or false), it is only necessary to examine all the white shapes which exist.

The stimuli (diagrams) consisted of four dots which could be connected by straight lines. Six rules were constructed which referred to the connections between these dots, and which involved the quantifiers, "every," "some," and "no." A previous investigation (Johnson-Laird, 1969) showed that doubly-quantified sentences differed in the ease with which they are understood. It was decided to use three levels of difficulty with two rules at each level.

1. *Doubly-quantified rules*
   (a) "Every dot is connected to some dot or other."
   (b) "No dot is connected to every dot."

2. *Conjunction or disjunction of doubly-quantified rules*
   (c) "There is a dot which is connected to a dot but no dot is connected to every dot."
   (d) "There is a dot which is not connected to any dot or every dot is connected to every dot."

3. *Triply-quantified rules*

(e) "There is a dot connected to a dot to which no other dot is connected."

(f) "Every dot is connected to a dot to which another dot is connected."

It was assumed that these rules increased in complexity over the three levels. Hence, it was predicted that the simpler the rule, the greater the insight which would be gained into the structure of the task. Such insight would be manifested in the absence of positive instances chosen by the subjects and the presence of negative instances.

Task

Each subject acted as his own control and tested the truth of the six rules. Their order of presentation was counterbalanced by using four 6 × 6 Williams squares. As the subjects selected each diagram they were, in addition, required to state whether it was a positive or a negative instance of the rule. They were told that positive instances had been placed inside the envelope and negative instances left outside the envelope. Mistakes in identifying the diagrams as positive or negative instances were corrected by the experimenter.

Materials

The same 11 diagrams were used throughout the experiment for all six rules. Each consisted of four circular dots ($\frac{1}{8}$ in. diameter), arranged in a square ($1\frac{1}{8}$ in. side), and between any two dots there either was, or was not, a straight line (see Table II). There

TABLE II

The eleven diagrams and the positive (1) *and negative* (0) *instances of the six rules*

	1	2	3	4	5	6	7	8	9	10	11
											
Rule (a)	0	0	0	0	1	1	1	1	1	1	1
Rule (b)	1	1	1	1	1	1	1	0	0	0	0
Rule (c)	0	1	1	1	1	1	1	0	0	0	0
Rule (d)	1	1	1	1	0	0	0	0	0	0	1
Rule (e)	0	1	1	0	1	0	1	1	1	0	0
Rule (f)	0	0	0	0	0	1	1	0	1	1	1

are $2^6 = 64$ possible diagrams which could have been used but they fell into 11 subsets in relation to the logic of the rules. One representative diagram was chosen from each of these subsets. (It will be noted that diagrams (2) and (3) are mutually redundant for the rules used in the present experiment. Other triply-quantified rules do, however, distinguish them.)

Each of the six rules was typed on a card (6 × 4 in.). The diagrams were drawn in heavy ink at one end of a card (5 × 3 in.) so that the subject could hold the card in his hand and see the diagram clearly.

Subjects

Twenty-four psychology undergraduates, who were native-speakers of English and unfamiliar with formal logic, were tested individually.

Procedure

The 11 diagrams were arranged in two rows, (1) to (6) and (7) to (11) in front of the subject. The experimenter explained that the task was not an intelligence test but a reasoning task. The subjects were told that their task was to prove whether or not each rule was true for the diagrams in the envelope. They were instructed to select those diagrams about which they needed information with respect to their location, and, as they handed them over, to say whether they were positive or negative instances of the rule. The experimenter then said whether each diagram was inside or outside the envelope. Although the subjects were told not to hurry, the interval during which they perused the rule prior to their first selection was covertly timed. It was stressed in both the verbal instructions and a written summary of them that even if a rule was true for the contents of the envelope, it might also apply to some of the diagrams outside the envelope.

Results

Of the 24 subjects, five showed no insight into the task—they consistently chose both positive and negative instances throughout the experiment. Four subjects showed complete insight into the task from the beginning. The data from these two sets of subjects were discarded as unilluminating.

The data from the remaining subjects were examined to discover the extent to which their insight into the task was a function of the complexity of the rules. Three qualitative types of performance were distinguished: trials on which only positive instances were examined, trials on which both positive and negative instances were examined, and trials on which only negative instances were examined. Mistakes in identification were left uncorrected for the purposes of this analysis: a subject has insight if he intends to choose a negative instance, and lacks it if he intends to choose a positive instance.

There was a slight but insignificant tendency for insight to increase with practice at the task. Table III shows the frequency with which the three types of performance occurred for each rule, together with the percentage error in identification of the instances as positive or negative and the mean inspection

TABLE III

Frequency of selecting instances, percentage error in identification for each rule, and mean inspection times (N = 15)

	Rules					
	(a)	(b)	(c)	(d)	(e)	(f)
Negative instances	9	13	10	2	7	9
Mixed instances	4	2	4	6	4	4
Positive instances	2	0	1	7	4	2
Percentage error in identification	1·5	1·9	2·3	1·5	3·0	7·2
Mean inspection times (sec.)	29	40	37	55	41	53

times. It is evident that the predicted trend is not confirmed because of the striking lack of insight for the disjunctive rule (d). However, the 15 subjects did show agreement with one another ($P<0{\cdot}01$), and the degree of insight exhibited for rules (a), (b) and (c) was significantly greater than for rules (d), (e) and (f) (Wilcoxon test, $P<0{\cdot}01$, one-tailed). In fact, only one subject went against the general trend. It will be noted that the error rate for identifying instances as positive or negative was, in general, extremely low. However, the case of understanding rules, indicated by the time subjects spent examining them prior to their first selection, does seem to be related to the amount of insight they had into the task. The mean times in Table III are a little misleading in this respect, since they are based on times from subjects with varying degrees of insight. The mean inspection time for rules associated with complete insight was 42 sec.; for rules associated with partial insight it was 50 sec.; and for rules associated with no insight it was 55 sec. Since many of the subjects performed at only two levels of insight, it was not possible to perform any simple trend test. But it was possible to perform a Sign test on the basis of whether the subject's mean inspection times (at each level of insight) correlated with the level of insight. For all but two subjects out of the 15, the tendency was for shorter inspection times to be associated with higher levels of insight ($P<0{\cdot}04$, two-tail).

The difficulty of the disjunctive rule was remarkable. For over half the subjects its introduction led to an apparent decrease in insight, and for only one subject was it associated with an increase in insight. It was unlike any other rule in this respect. Moreover, four subjects performed with apparent insight both on the two rules which occurred immediately before the disjunctive rule and on the rule which immediately followed it. But the disjunctive rule was tested by these subjects solely with positive instances. One of them said afterwards, "I don't know why I did that—it was silly." It is particularly important to note that the difficulty of the disjunctive rule is *not* due, as might have been expected, to the difficulty of distinguishing its positive and negative instances. Together with rule (a) it shows the lowest error identification rate (1·5 per cent) of all the rules.

Subjects failed to examine *all* the negative instances on exactly half of those occasions where they selected only negative instances. This might indicate only a partial insight into the task. A closer examination of the data, however, suggests that subjects merely failed to identify the complete set of negative instances. Consider, for example, what happened with rule (f): nine subjects examined only negative instances but only two of them examined the complete set. The remaining seven subjects all failed to examine stimulus (8). It thus seems that they had grasped the structure of the task but had failed to grasp this particular rule.

Introspective reports

Remarks made by the subjects both during and after the experiment were revealing. At least one subject attempted to translate the rules into visual images of the "characteristic diagram." Thus rule (b) yielded a "mental picture" similar to diagram (11). To some extent it was possible to keep track of the

level of insight by observing spontaneous remarks. "There's something I've got to crack here." One subject who had chosen all the positive and then all the negative instances of a rule remarked, "if I'd done it in the opposite order, I could have done it in four!" Another said "one positive diagram might be necessary to make sure that there *was* something in the envelope." This remark is perhaps analogous to the behaviour of the subjects in Experiment I who were assumed to draw stimuli immediately from the black box in order to establish the existence of a triangle.

Discussion

The results confirmed the main prediction that insight into the structure of the task would be affected by the nature of the rules. This confirmation occurred despite the relative failure of the predicted trend in the difficulty of the rules. It seems that in considering the "complexity" of a rule it is necessary to distinguish two separate factors: (1) the ease of understanding what the rule means, which presumably is reflected by the time spent examining it; and (2) the ease of identifying positive and negative instances of it, which is reflected in the number of identification errors.

It is evident that the second factor concerns the reference of the rule and is due to an interaction between its meaning and the nature of the connections in the diagrams. This difficulty of relating rules to diagrams came to a head with rule (f), "Every dot is connected to a dot to which another dot is connected." There were relatively many identification errors, and a particularly persistent one was to identify diagram (7) as a negative instance. As one subject put it, "I was thinking one way along a line instead of thinking both ways. In other words you must think both ways from each dot." This would seem to be a rather interesting example of an apparently irreversible thought process—a characteristic which, according to Piaget and his associates, is overcome in early childhood.

The ease of understanding the rules seems to have had a much greater effect upon performance. For example, subjects complained that "double statements," "negatives" and "ambiguity" caused difficulty. Ambiguity was detected in the phrase, "another dot." Did it mean just one other dot or possibly more than one? Similarly, one subject suggested that rule (c), "There is a dot connected to a dot to which no other dot is connected," might be better formulated as, "There is at least one dot which has only one connection." This is certainly easier to understand and, perhaps, a more conventional way of expressing the rule. The correlation between inspection times and level of insight suggests that these, and other, aspects of the process of understanding a rule influenced the amount of insight subjects had into the task.

The unexpected difficulty of the disjunctive rule seems to have been due, in part, to the problem of understanding what it meant: its mean inspection time is longer than that of any other rule (cf. Table III). It also seems to have been due to the difficulty of having to hold two mutually incompatible descriptions in mind at the same time. As one subject put it, "There were two types of diagram in the envelope." It is possible that this occupies a greater amount of short-term

memory than a single complex rule, and thus leaves a smaller amount of "computing space" available for handling the selection of the diagrams.

There was some support for Wason's (1961) observation concerning the way negative sentences are evaluated. The subjects were often initially confused about which were the positive instances and which were the negative ones, having correctly partitioned the set of diagrams. As Wason suggests, negatives may be mentally deleted, the resulting affirmative sentence evaluated, and the response inverted to take account of the negative. But this last stage could be easily forgotten, or overlooked. (This would be analogous to asking the question, "what is the opposite of 'not inside'?", and getting the answer, "outside.") But it is important to note that such a process, if it occurred, did not lead to overt errors in identification of the instances. The subjects merely seemed confused, and reported confusion before identifying the instances and requesting information about their location. Inspection of Table III shows that the error rate for the 15 relevant subjects is very low indeed.

But the most important result in this experiment is that insight into the implicative structure of the task may be temporarily lost as a function of the complexity of the material.

General Discussion

What is so difficult about the logical relation of implication? The previous experiments (Wason, 1968) suggest that it can be made formidably difficult. When the task is abstract, and when inferences have to be made about conditional sentences, then an extremely deep and erroneous fixation is likely to occur. But the present experiments show convincingly that implication can be made relatively easy to appreciate.

In Experiment I all the subjects did eventually gain spontaneous insight into the task. Given the sentence, "all the triangles are black," or "if they are triangles, then they are black," all the subjects eventually stop requesting black stimuli, and start requesting white ones. And this tendency is particularly pronounced when the instructions specify that the sentence has to be proved false rather than true, i.e. "some triangles are white." A set for truth naturally encourages the tendency to search for black triangles, and a set for falsity naturally encourages a search for white triangles. The results suggest that initially insight may be partially gained—some subjects drew out black and white stimuli alternately. But there is a definite stage when insight is apparently completely gained. No subjects asked for more than nine out of a possible 15 black stimuli, and they all exhausted the 15 white stimuli. Thus an optimistic generalization would be that familiarity with a simple task enables insight into the logical structure of implication to be permanently gained.

Such optimism is shown to be unfounded by the results of Experiment II. If only relatively simple rules had been used as material in this experiment, then it would have been inferred that insight into the structure of the task had been spontaneously and permanently gained, in just the same way as it seemed to have been gained in Experiment I. But the results show that when the experimental

material is complex, and particularly when rules involving exclusive disjunction are used, insight may be lost, even though it seemed to have been gained under simpler rules. And it is particularly important to stress that this is not just a matter of what might be assumed to be difficult. It is not a difficulty of correctly identifying positive and negative instances. It is a difficulty of realizing, under these conditions, that only negative instances are informative. This result is both disturbing and surprising. It goes against the traditional Gestalt conception of insight as an all-or-none phenomenon, and is perhaps more comparable with the effects of attempting to teach young children conservation, when a relatively trivial modification in the situation may lead to the evaporation of apparent insight.

This raises the question about whether the objective criteria for insight in the present experiments do really indicate insight. This hinges, of course, on the way in which insight is defined. It could be argued that if insight is so labile, then it is more apparent than real—it is not "really" insight at all. But this hardly does justice to the subjects' opinion. They described having had the insight but making the wrong response for reasons that escaped them. Hence it seems more reasonable to suggest that they were distracted from their appreciation of the structure of the task by the irrevelant complexity of the rules. All the evidence from the previous experiments suggests that implication is a difficult concept to grasp fully, and perhaps when it is grasped it is difficult to hold. When the subject is confronted by too much noise, in the form of irrelevant complexity, he may well fall back on the more primitive "matching responses" which were observed to occur frequently when implication was expressed in disjunctive form, "either not-p or q" (Wason and Johnson-Laird, 1969). In Experiment II "matching responses" would, of course, result in selection of only positive instances. Implication critically involves the appreciation of negation, and negation is a second-order, a more sophisticated concept than affirmation, in which a simple match exists between perceptual data and logical relation. If noise had been introduced in Experiment I, it would be predicted that the subjects would abandon their inspection of white stimuli and start requesting black stimuli again.

If, in fact, insight can be apparently gained, and then subsequently lost, the result is of considerable theoretical importance. In the past, attention has been concentrated on the conditions which facilitate the acquisition of insight, e.g. prior familiarity. More attention should be paid to the conditions which result in the disappearance of insight after it has been gained. The issue is not simply an academic one. It is relevant, for instance, to the construction of intelligence test items. A problem is usually solved when its logical structure is clear. What extraneous factors militate for, and against, insight into their structure?

These experiments show that the logical relation of implication does cause difficulty regardless of the way in which it is realized, i.e. as a conditional or quantified sentence, or even when it is not overtly expressed in a sentence, but is a structural feature in a reasoning task. Spontaneous insight is apparently gained into it, but such insight would appear to be often lost when the cognitive load is too great.

Grateful acknowledgements are due to Mrs. Diana Shapiro who prepared the stimulus material, and organized the recruitment of the subjects, and to the Medical Research Council for a grant for scientific assistance.

References

JOHNSON-LAIRD, P. N. (1969). On understanding logically complex sentences. *Q. Jl. exp. Psychol.* **21**, 1–13.

JOHNSON-LAIRD, P. N. and TAGART, J. (1969). How implication is understood. *Am. Jl. Psychol.* **82**, 367–73.

WASON, P. C. (1961). Response to affirmative and negative binary statements. *Br. Jl. Psychol.* **52**, 133–42.

WASON, P. C. (1968). Reasoning about a rule. *Q. Jl. exp. Psychol.* **20**, 273–81.

WASON, P. C. and JOHNSON-LAIRD, P. N. (1969). Proving a disjunctive rule. *Q. Jl. exp. Psychol.* **21**, 14–20.

WASON, P. C. (1969). Regression in reasoning? *Br. Jl. Psychol.* **60**, 471–80.

Received 6 *November* 1969