Psychol. Res. 41, 79–90 (1979)

Psychological Research
© by Springer-Verlag · 1979

THOG: The Anatomy of a Problem

Peter C. Wason[1] and Philip G. Brooks[2]

[1] Psycholinguistics Research Unit, University College London, England
[2] School of Behavioural and Social Science, Plymouth Polytechnic, Plymouth, England

Summary. Three experiments are reported on the attempts to solve a novel hypothetico-deductive problem. Its solution demands both the postulation of hypotheses about its structure and a combinatorial analysis upon the consequences of these hypotheses. The majority of subjects (students) failed to solve the problem because they argued from the properties of stimuli rather than from hypotheses about their conceptual status. The results suggest that a familiarity with the logical structure of the problem and the elicitation of appropriate hypotheses failed to correct this intuitive approach. These findings are discussed in relation to Piaget's theory of formal operations, and (very tentatively) in relation to habitual styles of thought.

Introduction

This paper is about an exploratory investigation of the way in which individuals tackle a novel hypothetico-deductive problem. The solution to such problems entails the postulation of hypotheses followed by a combinatorial analysis of their potential consequences.

The problems used by psychologists to study thinking have varied widely in the way in which they allow both these processes to occur. In the Gestalt tradition, Duncker's (1945) classic X-ray problem stimulates the subject to entertain a variety of hypotheses, but because the variables are not predetermined, it affords little scope for a combinatorial analysis in which 'the possible will be exhaustively inventoried' (Flavell, 1963). On the other hand, the realistic tasks used by Inhelder and Piaget (1958) to study adolescents' reasoning do allow for the possibility of combinatorial analysis, but we have argued they do not exhibit the exercise of formal operations unadulterated by the content in which they are presented (Wason and Johnson-Laird, 1972). Abstract formal problems are also open to criticism. It is becoming something of a fashion to denigrate them (e.g., Smedslund, 1977) on the grounds that they demand only uncreative, algorithmic procedures remote from the problems of everyday life. However, in response to such criticism, we have argued (Wason, 1978) that the artificial problems

0340–0727/79/0041/0079/$ 02.40

used by the psychologist are not intended to copy real-life situations. They are intended to capture some of the same processes, the puzzlement, the doubts, the obsessive tendencies towards repetition, which occur conspicuously during the crises of everyday life. Abstract hypothetico-deductive problems may reveal some of these processes when the individual fails to grasp their formal structure. It is now well known that formal operations may not be elicited by the presentation of a formal problem (e.g. Wason, 1977a). Indeed, Evans (e.g., Evans and Lynch, 1973) has adduced evidence that students resort to a primitive strategy of 'matching' variables when confronted by a difficult and novel problem; instead of systematically testing combinations of variables, they seek to achieve a one-to-one correspondence between them. Such a strategy is distinctively relevant to the present problem. The problem is designed to be such that, if the individual fails to carry out a combinatorial analysis based on the problem's structure, he reaches a solution which is almost the mirror image of the correct solution.

The Problem. The THOG problem (Wason, 1977b) has been formulated in several ways, but extensive pilot testing has shown that the following formulation appears to be the clearest:

'In front of your are four designs: Blue Diamond, Red Diamond, Blue Circle and Red Circle (see Fig. 1).

You are to assume that I have written down one of the colours (blue or red) and one of the shapes (diamond or circle.) Now read the following rule carefully:

If, and only if, any of the designs includes either the colour I have written down, or the shape I have writen down, but not both, then it is called a THOG.

I will tell you that the Blue Diamond is a THOG.

Each of the designs can now be classified into one of the following categories:

A) Definitely is a THOG.

B) Insufficient information to decide.

C) Definitely is not a THOG'.

The Solution. The problem is surprisingly difficult for many people although once the necessity of a combinatorial analysis is apparent it may seem trivial. The pseudo-solution is based on properties of the designs rather than on hypotheses. It goes something like this:

The Red Diamond and the Blue Circle each have one property in common with the Blue Diamond so each of these designs could be THOGS: there is insufficient information to be certain. The Red Circle has no property in common with the Blue Diamond so it definitely is not a THOG.

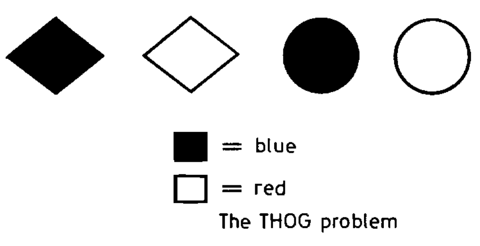

Fig. 1. The four designs used in the standard problem

The correct solution is that the Red Circle definitely is a THOG and that the Red Diamond and the Blue Circle definitely are not THOGS. The proof reflects the hypothetico-deductive nature of the problem. First consider all the possible combinations of features:

diamond and blue
circle and blue
diamond and red
circle and red

'Diamond and blue' could not have been written down because the Blue Diamond includes *both* these features, and this is disallowed by the rule. Similarly, 'circle and red' could not have been written down because the Blue Diamond includes *neither* of these features. But the two remaining hypotheses, 'circle and blue' and 'diamond and red,' each include *one* feature in common with the Blue Diamond, and hence could have been written down.

The combinatorial analysis follows. Consider first the 'circle and blue' hypothesis. If it were the correct one, the Blue Circle and the Red Diamond could not be THOGS because these designs respectively include *both* and *neither* of the features enumerated in the hypothesis. But the Red Circle *could be* a THOG because it includes 'circle' enumerated in the hypothesis. So far the outcomes are indeterminate because the other candidate hypothesis 'diamond and red', has not been considered. But this hypothesis leads to the same result. The Blue Circle and the Red Diamond could not be THOGS because these designs include respectively *neither* and *both* of the features enumerated in the hypothesis. But the Red Circle *could be* a THOG because it includes 'red,' enumerated in the hypothesis. So whichever hypothesis is the correct one the outcomes agree, but of course both have to be tested to prove the conclusion.

The entire proof can be short-circuited into a couple of lines, and subjects sometimes reason in this way. From a consideration that the Blue Diamond is a THOG, the two hypotheses may be expressed as follows:

1) If the shape (written down) is diamond, then the colour (written down) is red.
2) If the colour (written down) is blue, then the shape (written down) is circle.

Reading across each line, it follows that the Red Diamond and the Blue Circle are excluded, and reading vertically between the lines, it follows that the Red Circle is included in the same way that the Blue Diamond is included.

Previous Research. Some useful preliminary work has been carried out on the problem by Cordell (1978) using group testing procedures on large samples. The results showed that modifications of the problem, in the attempt to make it easier, had no beneficial effects whatsoever. These included an 'anti-coalescense' presentation of the stimuli, i.e., the two attributes in each design were discrete stimuli rather than conjoined (a number and a letter rather than a coloured shape), and an attempt to thematize the problem, i.e., presentation of the material in a realistic guise. Interestingly enough, this latter mode of presentation made the problem (if anything) more difficult.

Mimikos (personal communication) has demonstrated in a very surprising study that students with a Science background perform very much better on the problem than students with an Arts background. However, there are strong grounds for supposing that the Science sample was unduly weighted with mathematical and computational skills.

The original (Wason, 1977b) formulation of the problem is now considered to be too concise. It was: 'In the above designs there is a particular shape and a particular colour such that any of the four designs which has one, and only one, of these features is called a THOG.' This differs in two ways from the present formulation. First, it does not seem to create a sufficient demarcation between the hypotheses and the designs; second, the attempt to avoid the negative in the exclusive disjunction by the expression 'one, and only one, of these features,' runs the risk of being positively misleading. Pilot studies suggested that a few subjects equated this expression, not with particular features but with any features at all. Thus, contrary to the designated positive instance, they assumed a THOG would possess either a colour (alone) or a shape (alone) – a kind of disembodied THOG. Cordell retained this expression, but introduced the present strong demarcation between the hypotheses and the designs due to E.A. Lunzer (personal communication).

The Present Investigation. This research aims to determine whether the pseudo-solution to the THOG problem can be inhibited by a) acquainting the subjects with the structure of the problem, and b) partitioning the problem into two parts corresponsing to a formulation of hypotheses and to the combinatorial analysis.

Experiment 1

Suppose someone were made familiar with the structure of the problem. Suppose (more accurately) that someone were allowed to construct the problem and discover for himself that two instances of THOG necessarily had no property in common. Would this acquaintance with the bones of the problem enable him to solve it *ab initio*? This discovery of the problem's structure amounts to knowing those features which the Experimenter had written down. It is, of course, not the same as solving the problem because in the standard problem the initially given instance is consistent with one out of two pairs of features predetermined by the Experimenter, and the subject cannot know which pair is the correct one. It follows that if the construction of the problem in terms of fixed features is unhelpful, then the crucial difficulty of the problem would seem to reside in its hypothetical nature rather than in its logical structure.

Design. An experimental and a control group were used to test the hypothesis that there would be positive transfer from constructing the problem to solving the standard problem. Both groups attempted two problems, the second one being the identical THOG problem. The first problem was isomorphic to THOG but used different stimulus material (Green Square, Brown Square, Green Triangle, Brown Triangle), and a different designated name: CHUZ. The difference between the experimental and control groups resided in the presentation of this problem.

In the experimental group, after presenting the four designs, the subjects were instructed:

'Please write down one of the colours (green or brown) and one of the shapes (square or triangle) in the space provided. If, and only if, any of the designs includes either the colour which you have written down, or the shape which you have written down, but not both, then it is called a CHUZ.'

The task was to assign each of the designs into one of these categories: A) Definitely is a CHUZ; B) Insufficient information to decide; C) Definitely is not a CHUZ; and to provide reasons for the assignments.

In the control group the instructions were the same as those for the standard THOG problem, and an instance (Green Square) was designated a CHUZ. Thus the CHUZ problem in the control group differed from the standard THOG problem solely with respect to the material and the name.

After completion of the initial CHUZ problems both groups were presented with the THOG problem. The following instructions preceded the initial problem: 'Please read through this page carefully and ask any questions before you begin writing. Take your time since the problem is more difficult than it might seem.' Performance on both problems was covertly timed.

Subjects. Twenty-eight paid volunteer undergraduates of Plymouth Polytechnic were allocated alternately to the groups and tested individually.

Results. Tables 1 and 2 show the number of subjects who were correct and incorrect on the two problems for the experimental and control groups respectively.

It is apparent from Table 1 that all the subjects in the experimental group were correct in their construction of the CHUZ problem, and yet only 4 out of 14 were correct on the subsequent THOG problem. Inspection of Table 2 reveals a similar pattern of performance on the THOG problem in the control group: five subjects were correct. There is clearly no facilitation from constructing the problem, and the prediction is thereby not confirmed.

Table 1. Relation between solving the constructed CHUZ problem and the standard THOG problem in the experimental group (Experiment 1)

		Constructed CHUZ Problem		
		Correct	Incorrect	Total
Standard THOG Problem	Correct	4	0	4
	Incorrect	10	0	10
	Total	14	0	14

Table 2. Relation between solving the standard CHUZ problem and the standard THOG problem in the control group (Experiment 1)

		Standard CHUZ Problem		
		Correct	Incorrect	Total
Standard THOG Problem	Correct	3	2	5
	Incorrect	0	9	9
	Total	3	11	14

The responses to the THOG problem (regardless of group) provide a conservative index of the subjects' thought processes. Out of the erroneous solutions, the largest class was what we have called the 'pseudo-solution' (Blue Circle and Red Diamond indeterminate and Red Circle definitely not a THOG). Eight out of 19 errors fell into this class. A closely related error (Blue Circle and Red Diamond are THOGS and Red Circle definitely not a THOG) occurred once. The former error is called 'matching' and the latter 'inverse' by Cordell (1978). We shall conflate them, and refer to them as 'intuitive errors' because both seem due to a plausible inference based on the properties of the designs rather than on the hypotheses. Of the remaining errors, 5 consisted of 'insufficient information' to all unspecified designs, 2 exhibited 'near insight' (insufficient information to decide about Blue Circle and Red Diamond but Red Circle is a THOG), and 3 appeared arbitrary. If the classification of Blue Diamond is discounted (since it is given as a THOG), then the chance probability of any solution is 1 in 27 (.037). It follows that the responses are far from random, 9 out of 19 (47.37 %) of erroneous solutions exhibit 'intuitive error'.

The latencies to perform the tasks were unilluminating. As might be expected, the mean time taken to construct the CHUZ problem (6.85 min) was significantly much less than the mean time taken to solve it in the control group (13.19 min). The corresponding times taken to solve the THOG problem were 8.77 min and 7.32 min, in the experimental and control groups respectively. These times did not differ significantly.

It would seem from the results of this experiment that discovering the structure of the problem – an awareness of the correct classification of the designs under the rule – confers no advantage whatsoever on the ability to solve the problem. It would accordingly seem that either the ability to postulate the appropriate hypotheses, or the ability to conduct a combinatorial analysis on their outcomes, is responsible for difficulty. Experiment 2 sought to elucidate this aspect of the problem.

Experiment 2

This experiment aimed to determine whether the elicitation of the hypotheses underlying the solution to the problem would facilitate its solution, and whether such facilitation would be generalized to a different version of the same problem. 1) Subjects were presented with the materials of the CHUZ problem (hereafter called the 'partitioned CHUZ problem') and a designated positive instance: Green Square. But they did not at this stage have to solve the problem. 2) They were asked to carry out an operation called 'identification of hypotheses' which consisted in placing a tick (and a cross) against those shape and colour combination(s) which the Experimenter could (and could not) have written down consistent with the given positive instance. The combinations were presented as follows:

triangle and brown . . . .
triangle and green . . . .
square and brown . . . .
square and green . . . .

3) Subjects giving correct answers at this stage were counted as 'spontaneously correct' and proceeded to the next stage. Subjects who gave the wrong answer were told the correct answer ('triangle and green' and 'square and brown'), and the reasons for this

were discussed until they claimed to have understood. Such subjects were counted as 'spontaneously incorrect.' 4) All subjects then proceeded to solve the partitioned CHUZ problem; if they succeeded in so doing they were scored as 'correct' in the rows of Table 3. No knowledge of results was given about the correctness of the solution. 5) Finally, all subjects attempted the standard THOG problem.

It was predicted that performance on THOG would be superior to performance on the same problem in the control group of Experiment 1. A similar 'extended treatment' was tested by Cordell (1978) without success, but he made no attempt to correct any erroneous identification of hypotheses.

Subjects. Fourteen paid volunteer undergraduates of Plymouth Polytechnic were tested individually.

Results. Table 3 shows the relation between spontaneous correctness on the identification of hypotheses and performance on the CHUZ problem.

This result is surprising. Although 64.3% (9 out of 14) of the subjects were spontaneously correct in identifying the relevant hypotheses, only 33.3% (3 out of 9) evidently conducted a combinatorial analysis upon them to achieve the solution. Similar results were obtained by Cordell (1978) – more than twice as many subjects were correct in the hypothesis identification stage, and then failed on the problem compared with those who were correct at both stages. Furthermore, no subjects benefited from the corrective treatment; all five who were corrected failed on the problem. There was no consistent pattern in the hypothesis identification for these subjects. One claimed that all hypotheses were tenable; one that none was tenable; two that all hypotheses were tenable other than 'triangle and brown'; one that all were tenable other than 'triangle and green.'

Table 4 shows the relation between performance on the CHUZ problem and on the THOG problem. It is evident that there is only a small (and non-significant) improvement between CHUZ and THOG. Similarly, comparison between performance on THOG in this experiment and performance on THOG in the control group of Experiment 1 reveals an identical distribution of correct and incorrect responses.

Of the errors made on THOG, 6 out of 9 fell into the category of 'intuitive error,' 1 consisted of 'insufficient information' throughout, and 2 appeared arbitrary. Thus 66.67% were 'intuitive', an increase over the 47.37% observed in Experiment 1. And these errors recurred in spite of the fact that all subjects had elicited (or been forced to entertain) the relevant hypotheses underlying the solution to the CHUZ problem.

Table 3. Relation between identification of hypotheses and solving the partitioned CHUZ problem (Experiment 2)

		Identification of hypotheses		
		Spontaneously Correct	Spontaneously Incorrect	Total
Solution	Correct	3	0	3
of	Incorrect	6	5	11
Problem				
	Total	9	5	14

Table 4. Relation between solving the partitioned CHUZ problem and the standard THOG problem (Experiment 2)

		Partitioned CHUZ Problem		
		Correct	Incorrect	Total
Standard	Correct	2	3	5
THOG	Incorrect	1	8	9
Problem				
	Total	3	11	14

It can be inferred from these results that *if* subjects are required to elicit the relevant hypotheses, then the majority can do so but only a minority of these subjects then proceed to solve the problem correctly. Thus the necessary completion of the combinatorial analysis is evidently more difficult than a consideration of the relevant hypotheses. The computational aspects of the problem appear more difficult than a grasp of its structure. What the identification of hypotheses does indicate, however, is that the majority of subjects do not *misunderstand* the problem.

However, there is a flaw in this experiment. The provision of a check-list of possible hypotheses may have exerted too explicit a constraint on the subjects' thought processes. Such a technique rules out the elicitation of a hypothesis consistent with 'intuitive error,' and so one cannot observe whether the subjects are solving a different problem to the one posed for them. The third experiment sought to remedy this defect by allowing for a more natural and spontaneous elicitation of hypotheses.

Experiment 3

In this experiment the subjects were not required to solve the problem but merely to elicit the relevant hypotheses. The CHUZ problem was used: Green Square, Brown Square, Green Triangle, Brown Triangle, with Green Square given as an instance of CHUZ. The 'intuitive' solution is that there is insufficient information to decide about Brown Square and Green Triangle, but the Brown Triangle is definitely not a CHUZ. Now how might a subject interpret the problem in order that this solution is entailed? Clearly, he would have to assume that the Experimenter had written down 'green' *or* had written down 'square.' The experiment sought to investigate the incidence of such responses, as well as seeing whether the results of Experiment 2, i.e., a predominance of correct hypotheses, would be elicited in a more spontaneous situation.

Design. The CHUZ problem was presented, but instead of providing a check-list of all possible hypotheses, subjects were instructed to record their answer to the following question: 'Your task is to determine, given the above information, what I could have written down, and to state your reasons briefly.'

Subjects. Ten paid volunteer undergraduates of Plymouth Polytechnic were tested individually.

Results. Eight out of the 10 subjects wrote down the correct hypotheses ('triangle and green' and 'square and brown'), six of them with adequate reasons. One subject wrote down all possible hypotheses, and one was unable to suggest any combinations at all although he struggled for about 30 min with the problem. Contrary to expectation, not a single subject recorded the hypothesis consistent with 'intuitive error' which we have seen is the predominant erroneous category.

These results, obtained in a freer situation, thus replicate those of Experiment 2, and suggest that the provision of a check-list of possible hypotheses does not act as an artefactual reminder. Combining the results from the two experiments, it is apparent that 70.8% of the subjects can elicit the relevant hypotheses, and thus grasp the logical structure of the problem.

Discussion

On a conservative estimate the prevalence of what we have called 'intuitive error' is a distinguishing feature of the results. Of the errors made, 53.37% fall into this category when the second (THOG) problem is attempted in spite of the fact that the chance probability of any error is 1 in 27. Thus there is a pronounced tendency to make inferences about the designs from instantiated properties ('blue' and 'diamond') and to ignore non-instantiated properties. This makes the Blue Circle and the Red Diamond potential candidates for THOGNESS, and it rules the Red Circle out of court.

However, the results of Experiments 2 and 3 suggest that a substantial majority of subjects have understood the logic of exclusive disjunction which underlies the THOG rule. It is evidently not the case that (on the whole) this rule is too complex to be grasped. In Experiment 2 nine out of 14 subjects spontaneously identified correctly the two hypotheses consistent with the positive instance, and in Experiment 3 eight out of 10 stated them when they were not prompted by a list of possibilities. Thus 70.8% of the subjects do accept the correct premises, and yet (in Experiment 2) 66.7% fail to solve the problem. This failure rate is increased to 78.6% if the subjects who are corrected are included. This result argues strongly against Henle's (1962) claim that mistakes in reasoning never occur but are always due to a misinterpretation of the premises. In the present problem we have tested an understanding of the premises independently of the solution.

But, of course, it cannot be inferred from the results of these experiments that the combinatorial analysis is the sole stumbling-block to the solution of the standard problem. It is evidently not the case that in attempting to solve the standard THOG problem, the subjects first elicit the relevant hypotheses, and then relapse. On the contrary, 'intuitive error' characterizes a completely different approach to the problem. What the results of Experiments 2 and 3 do indicate is that under the influence of experimental intervention the subjects do tend to meet a test of understanding satisfactorily, but that this test seems a remote and alien activity from the actual solution of the problem. In particular, the elicitation of hypotheses tends not to spur the combinatorial analysis. This may seem odd in relation to the results of Experiment 1.

For Experiment 1 shows conclusively that all the subjects solved the problem when they were allowed to 'fix' the hypotheses by writing down themselves a particular colour and a particular shape. In this way the subjects in the experimental group constructed the problem, and were able to mimic a combinatorial analysis based on the two features they had selected. This knowledge if gained at all, did not transfer to the solution of the standard THOG problem. One obvious difference, of course, is that the consequences of two pairs of features have to be tested in THOG as opposed to one pair in the constructed CHUZ. A deeper difference is that the determination of the critical features in the constructed CHUZ renders them necessarily not hypothetical at all. They become data rather than hypotheses. In the standard THOG problem, and in the partitioned CHUZ problem used in Experiment 2, the elicited hypotheses are real hypotheses because it is not known which one corresponds to reality. It may be peculiarly difficult to reason about hypothetical possibilities to reach determinate conclusions.

Inhelder and Piaget (1958) have shown that hypothetico-deductive thinking (formal operations) does appear to be mobilized by special kinds of task, the responses to which lie within the subjects' repertoire. It is the content of such tasks, e.g., chemical analysis, which may induce an appropriate strategy. In abstract tasks, without the possible benefit of thematic support, individuals may rely on matching the properties of the given instance with the remaining designs. In ordinary life if something shares a property with something else then it seems that they may come under the same class, and if two instances share no property in common then it seems they must come under different classes. Thus a strategy appropriate to the THOG problem may become dominated by a more routine strategy when the designs of the problem actually have to be categorised. In this way 'intuitive error' arises.

The results provide additional evidence against Piaget's theory of formal operations, at least in its strong form (Beth and Piaget, 1966). Just as in the results on Wason's 'selection task' (Wason, 1968), which involves the selection of the necessary and sufficient information to test the truth of a conditional sentence, the subjects in the present experiments conspicuously fail to carry out a combinatorial analysis on the variables of the problem. They tend to be driven by what is perceptually present rather than by systematically taking account of the possible in an exhaustive manner. In fact the present results may be more convincing because the THOG problem is not nearly so difficult as the selection task. When attempted second, a third of the subjects succeeded, and nearly all acknowledged the correctness of the solution subsequently. In contrast, the estimated success rate on the selection task (in 1970) was 3.9% and in a number of cases individuals were notoriously reluctant to accept the solution as correct. Cordell (1978) has even argued that his own results cast doubt on Piaget's (1972) weaker model of formal operational thought in which the necessity of posing the problem in a form familiar to the subject is acknowledged.

'Yet the THOG problems faced by our mature subjects would not, it may be argued, be all that alien: these students embarked upon degree or near-degree courses, will be used to conundrums in a variety of fields. Surely, then, with these formal reasoners one might have expected almost universal success in a problem couched in language no more specialized or recondite than that found in problems on the back of matchboxes?'

Can we say anything at all about those factors which allow one individual to solve THOG and another to succumb to 'intuitive error'? Obviously the present results must qualify those of Mimikos who found that students studying Science subjects did much better than those studying Arts subjects. Nearly all the students in the present samples had a background in science. And yet it may be that individual differences in cognitive functioning are operative. Chitra Kar (personal communication) has demonstrated an extremely strong association between performance on THOG and performance on the

'Embedded Figures Test' which is assumed to be an index of 'field dependence' (Witkin, 1965). Those subjects who were correct on THOG (10 out of 40) took about one-sixth of the time to extract a simple figure from a complex design compared to those who were incorrect on THOG. However, no differences were found between students with an arts and a science background. In spite of the doubts we have cast on Mimiko's extreme results, it would not be surprising to find differences in cognitive style reflected in adult specialisation. It may be that the habitual thinking of (say) art critics or aestheticians is associated with a more holistic mode of functioning, and that the habitual thinking of (say) mathematicians is associated with a more analytic mode of functioning. The results of two pilot studies on THOG provide grounds for speculation that this problem does differentiate two modes of thought.

In one of these studies a sample of 14 medical students and young doctors was tested. Seven solved the problem, taking an average of 6.3 min to do so. This is quite an impressive result. One young doctor (not in the sample), who solved the problem in his head in about one minute, pointed out that it was just like a diagnostic problem in which the symptom itself must not distract attention from its possible causes. He said 'I would not let any doctor near me who couldn't solve that problem.'

In the other study a sample of 10 student barristers was tested in an informal setting. All 10 failed to solve the problem. An attempt was made afterwards to explain its logic. One of the students argued for one hour against the correctness of the solution. Some of his observations, which tended to be highly repetitive, are worth quoting.

'I don't agree with you at all ... the Red Circle has nothing in common with the Blue Diamond. There is a fallacy... there must be. We can assume that either Blue or Diamond.. you can't have a circle – that is totally fallacious. Your argument does not stand up... either it is the diamond shape, or it must be the colour. You are right about the Blue Circle and the Red Diamond but I still think you are wrong about the Red Circle. Because if diamond shape is important... that is very clever... that is very clever indeed. It is a total reversal of all my thinking. You don't know which features are important – that is what it is all about. You seem to have two sets of features which are irreconcilable... at first sight you are given two unknown factors. But your mind has the unconscious assumption of one factor being definite. As soon as you get rid of the unconscious assumption, then you have worked things out. I find the logic very difficult to reconcile.'

Hypothetico-deductive thinking does form the basis of medical diagnosis, but it *may* not be an ingredient of legal argumentation which is presumably based on facts rather than possibilities. In this connection it is relevant to point out that in Mimikos' Arts sample 24 out of the 32 students were also students of Law. Not one of them solved the problem. However, these ancillary results are offered, not as systematic observation, but as speculative possibilities to stimulate further research on individual differences.

In one sense the THOG problem is utterly trivial. It has neither the surprising ingenuity of the best Gestalt problems, nor the conceptual interest of the 'selection task' with its emphasis on the distinction between verification and falsification. And yet it is non-trivial in the sense that it is astonishingly easy for some individuals and astonishingly difficult for others. At the very least it does seem a cogent test of formal operational thought in an abstract sphere. It sharply discriminates between two modes of attack. Individuals who solve it must postulate hypotheses and they must work out the consequences of these hypotheses in a combinatorial analysis.

Acknowledgements. This research was supported by a grant to the first author from the Science Research Council. We should like to thank Eric Lunzer and his students, especially Robert Cordell and Gordon Hyde-Dryden for stimulating discussions, and Jane Huges and Vera Marcus for arranging the medical and legal pilot studies.

References

Beth, E.W., Piaget, J.: Mathematical epistemology and psychology, Dordrech: Reidel 1966

Cordell, R.L.: Mature reasoning and problem solving. Unpublished. University of Nottingham M.Ed. dissertation (1978)

Duncker, K.: On problem solving. Psychol. Monogr. 58, whole No. 270, 1–113 (1945)

Evans, J.St.B.T., Lynch, J.S.: Matching bias in the selection task. Br. J. Psychol. 64, 391–397 (1973)

Flavell, J.H.: The developmental psychology of Jean Piaget. Princeton, N.J.: Van Nostrand 1963

Henle, M.: On the relation between logic and thinking. Psychol. Rev. 69, 366–378 (1962)

Inhelder, B., Piaget, J.: The growth of logical thinking from childhood to adolescence. New York: Basic Books 1958

Piaget, J.: Intellectual evolution from adolescence to adulthood. Hum. Dev. 15, 1–12 (1972)

Smedslund, J.: Piaget's psychology in practice. Br. J. Educ. Psychol. 47, 1–6 (1977)

Wason, P.C.: Reasoning about a rule. Q. J. Exp. Psychol. 47, 1–6 (1968)

Wason, P.C.: The theory of formal operations – a critique. In: Piaget and Knowing, Geber, B., ed., London: Routledge 1977a

Wason, P.C.: Self-contradictions. In: Thinking: readings in cognitive science, Johnson-Laird, P.N., and Wason, P.C., eds., Cambridge: Cambridge University Press 1977b

Wason, P.C.: Hypothesis testing and reasoning. Unit 25, Block 4, Cognitive psychology. Milton Keynes: Open University Press 1978

Wason, P.C., Johnson-Laird, P.N.: Psychology of reasoning: Structure and content. London: Batsford 1972

Witkin, E.A.: Psychological differentiation and forms of pathology. J. Abnorm. Psychol. 70, 317–336 (1965)

Received May 24 / September 19, 1979