

Rationality, memory and the search for counterexamples*

J.V. OAKHILL

University of Sussex

P.N. JOHNSON-LAIRD

MRC Applied Psychology Unit, Cambridge

Abstract

Two experiments investigated the failure to select potential counterexamples in testing generalisations. Subjects had to test whether a description of the contents of an envelope (a set of diagrams) was true or false. Since any diagram that fits the description could be inside or outside the envelope without affecting the status of the description, the rational strategy is to choose diagrams that do not fit the description since they can in principle falsify it. The subjects selected diagrams from a duplicate array in front of them, and the experimenter identified each selection as inside or outside the envelope. The first experiment established that disjunctive descriptions cause problems: subjects seldom gain an initial insight into the task with them, and sometimes lose an earlier insight. Their effects appear to be exacerbated by the presence of a negation in the description. The second experiment examined disjunctions and negations systematically, balancing the number of positive and negative instances of each description. It replicated the effects of disjunction, but not those of negation. The pattern of errors and latencies suggested a general explanation for failures in rationality: subjects do not see immediately the need to search for counterexamples, and may lose their insight into this principle when it becomes difficult to determine what counts as a counterexample.

*This research was supported by a grant from the Social Science Research Council to the authors. We would like to thank Alan Garnham for commenting on an earlier draft of the paper, and Julie Oakhill for preparing Tables 1 and 3. Reprint requests should be sent to J.V. Oakhill, MRC Perceptual and Cognitive Performance Unit, Laboratory of Experimental Psychology, University of Sussex, Brighton, BN1 9QG, United Kingdom.

Introduction

Counterexamples play a central role in rational thought. A generalisation is only true if there is no counterexample to it. Hence, an ability to search for counterexamples is a prerequisite for success in acquiring concepts, in developing and testing hypotheses, and in making and evaluating inferences. Previous research has shown, however, that subjects often fail to perform rationally and instead seek confirming evidence. This phenomenon has been observed in concept attainment (e.g. Bruner, Goodnow and Austin, 1956), in the discovery of hypotheses (e.g. Wason, 1960), and in deductive inference (e.g. Johnson-Laird and Steedman, 1978). But, the most striking occurrence of the phenomenon is in Wason's selection task, in which subjects have to test whether a general rule is true or false (see e.g. Evans, 1982; Wason and Johnson-Laird, 1972). Given the rule:

If there is a vowel on one side of a card, then there is an even number on the other side

the overwhelming majority of subjects choose to turn over a card bearing a vowel, but they do not select a card bearing an odd number even though the presence of a vowel on its other side would decisively falsify the rule.

Why do people fail to search for counterexamples? There appear to be two main possibilities. First, they may not appreciate the need to search for them. Second, they may grasp their relevance in principle, but fail to search for them properly. There can be few individuals who do not realise that the general assertion 'All swans are white' is falsified by the existence of a black swan. Yet, as performance in the selection task shows, subjects frequently do not realise the need to search for counterexamples: they examine swans to see if they are white, but pass over the opportunity to examine things that are not white to see if they are swans. In fact, with realistic materials of this sort, performance is often much better, though it is not invariably improved. Such is the lability of insight with realistic materials that some theorists propose a central role for memory: insight depends on the task triggering an appropriate memory for the content of the problem, the general rule, and its counterexamples (Manktelow and Evans, 1979). This claim appears to be too strong: experiments have shown that subjects must do more than merely remember the correct answer, since they can perform correctly with rules that they could never have encountered before. The most that can be safely claimed is that a memory for an *analogous* rule may suffice to produce insight (Griggs, 1983; Griggs and Cox, 1982).

There is quite a different factor that may adversely affect the search for counterexamples: the cognitive load of the task. This conjecture is suggested

by a neglected result of Johnson-Laird and Wason (1970). They used a reduced version of the selection task in which the subjects had to choose between just two classes of stimuli: those that were consistent with a general rule and those that were inconsistent with it. With this reduced array of stimuli, the subjects showed a much greater insight into the need to select potential counterexamples. In a further experiment, the complexity of the materials was manipulated to see whether it affected performance. The relation of implication was not stated in the materials, but was inherent in the structure of the task itself: the subjects had to test whether a *description* of the contents of an envelope was correct. The potential contents of the envelope consisted of a set of diagrams, and the subjects carried out the task by selecting diagrams from a duplicate array in front of them. As they chose each diagram, it was identified by the experimenter as either inside or outside the envelope. A diagram that fits the description could, of course, be inside or outside the envelope, and the description could still correctly describe the contents of the envelope. But, a diagram that does not fit the description must be outside the envelope, or else the description is false of the contents of the envelope. The subjects were offered the choice between these two sorts of stimuli. The correct strategy is, of course, to choose counterexamples, i.e. diagrams that do not fit the description, since only they can falsify it. The subjects usually did not appreciate the logical structure of the task at first, and they chose to ask about diagrams that were positive instances of the description. During the course of the experiment, most subjects realised that they had adopted an inappropriate strategy, and started to ask about only negative instances of the description. However, there was evidence that such apparent 'insight' was often temporarily lost, as a function of the complexity of the descriptions. The subjects found the one disjunctive description particularly troublesome, frequently reverting to their previous strategy of choosing positive instances of the rule. Johnson-Laird and Wason offered the post hoc speculation that its difficulty might be the result of the load on memory that a disjunction imposes, since it calls for two mutually incompatible descriptions to be held in mind at the same time, thus leaving a smaller amount of 'computing space' available for the selection of the diagrams.

People have to work out for themselves the need to search for counterexamples, and this discovery and its maintenance in practice may be impeded by the cognitive load of the task. Our first experiment was designed to follow up this idea, since Johnson-Laird and Wason reported the effect for just one disjunctive sentence. In order to manipulate cognitive load, we used three sorts of descriptions that varied in complexity: simple categorial descriptions, inclusive disjunctive descriptions (in which the sets of diagrams described by the two disjuncts overlap to some extent), and exclusive disjunctive

descriptions (in which the two disjuncts describe completely distinct sets of diagrams). We assumed that the disjunctive assertions would be more complex than the simple assertions, and also that exclusive disjunctions would be more complex than inclusive disjunctions. We predicted that complexity would affect insight into the logic of the task. In order to perform the task correctly, it is necessary to select counterexamples to the description of the envelope's contents. However, the more complex the description, the smaller the amount of available capacity for coping with the logic of the task. Hence, the number of selections of positive instances of the descriptions should increase with the complexity of the descriptions (from simple descriptions to exclusive disjunctions), and an initial insight into the implicative structure of the task should be more likely to occur with a simpler description.

Method

Subjects

Thirty members of the student and staff population of the University of Sussex were paid £1.50 to participate in the experiment, which lasted approximately 40 minutes.

Materials

The eleven diagrams were identical to those used by Johnson-Laird and Wason. Each of them consisted of four dots, and between any two dots there was sometimes a straight line and sometimes not. There were nine descriptions, each of which described a subset of the diagrams, with three rules at each level of complexity:

Simple descriptions

- 1a. Every dot is connected to at least one other dot.
- 1b. At least one dot is connected to all the dots.
- 1c. No dot is connected to all the dots.

Inclusive disjunctions

- 2a. Every dot is connected to at least one other dot,
OR at least one dot is connected to all the dots.
- 2b. No dot is connected to all the dots,
OR no dot is connected to any other dot.
- 2c. No dot is connected to two other dots,
OR no dot is connected to all the dots.

Exclusive disjunctions

- 3a. Every dot is connected to at least one other dot,
OR no dot is connected to any other dot.
- 3b. No dot is connected to any other dot,
OR at least one dot is connected to all the dots.
- 3c. At least one dot is connected to all the dots,
OR no dot is connected to two other dots.

Each rule was typed onto an index card. Table 1 shows the eleven diagrams, and the positive and negative instances of each description. The word 'connected', of course, refers to a direct link from one dot to another.

Procedure

The subjects were tested individually in a quiet room. They sat at a table on which the 11 diagrams were arranged. The subjects' instructions were similar to those used by Johnson-Laird and Wason (1970). Subjects were told that they should imagine that the experimenter had taken copies of some of the diagrams, put them in an envelope, and written a description of the set of

Table 1. *The eleven diagrams in Experiment 1, and the positive (1) and negative (0) instances of the nine descriptions*

	1	2	3	4	5	6	7	8	9	10	11
Descriptions											
1a	0	0	0	0	1	1	1	1	1	1	1
1b	0	0	0	0	0	0	0	1	1	1	1
1c	1	1	1	1	1	1	1	0	0	0	0
2a	0	0	0	0	1	1	1	1	1	1	1
2b	1	1	1	1	1	1	1	0	0	0	0
2c	1	1	1	1	1	1	1	0	0	0	0
3a	1	0	0	0	1	1	1	1	1	1	1
3b	1	0	0	0	0	0	0	1	1	1	1
3c	1	1	0	0	1	0	0	1	1	1	1

diagrams it contained. They were given an example description, and it was explained that, for the description to be true of the contents of the envelope, all diagrams in the envelope must fit that description. The verbal instructions and a subsequent written summary stressed that not *all* the cards fitting the description needed to be in the envelope for it to be true, and that there was always at least one card in the envelope. The subjects' task was to check whether each description was true of the contents of the envelope, and they were told to ask for information about the whereabouts of those cards that they felt were relevant to establishing the truth or falsity of each of the rules. It was explained that the disjunctions should be taken to mean A or B or both, rather than as expressing mutually exclusive alternatives, since the latter interpretation would imply that diagrams conforming to both parts of a disjunctive description could not be inside the envelope. The experimenter also explained that the phrase 'connected' referred to a direct connection from one dot to another, and checked that subjects understood this point. Finally, the subjects were told that they should aim at establishing their conclusions using as few diagrams as possible for each description.

Each subject received all nine descriptions in a random order. The time between presentation and the subjects' selection of the first diagram was measured for each description, using a hidden timer activated by a foot switch.

On selecting a diagram, the subjects had to state whether it fitted the rule or not, and any false identifications were corrected by the experimenter. Positive instances were always confirmed to be inside the envelope, and negative instances, outside. This procedure was continued for each description in turn, until the subject was satisfied that the description was true of the contents of the envelope.

Results

Of the 30 subjects tested, 8 showed no insight into the task, selecting both positive and negative instances throughout the experiment, and a further 3 subjects showed complete insight from the beginning. The data from these 11 subjects were discarded from the analyses. The data from the 19 remaining subjects were divided into three categories: trial on which only positive instances were chosen, trials on which both positive and negative instances were chosen, and trials on which only negative instances were chosen.

Table 2 shows the frequency with which the three types of responses occurred for each description, together with the number of errors in identification of instances of each description, and the mean inspection times.

Table 2. *The number of subjects in Experiment 1 selecting negative, mixed, and positive instances for each of the nine descriptions; the mean inspection times prior to the first selection; and the total number of errors in identifying instances of the descriptions.*

	Rules								
	Categorical			Inclusive disjunctions			Exclusive disjunctions		
	1a	1b	1c	2a	2b	2c	3a	3b	3c
Negative instances	16	8	13	14	10	10	10	8	10
Mixed instances	3	8	6	5	9	8	9	9	8
Positive instances	0	3	0	0	0	1	0	2	1
Inspection times (s)	30	35	33	53	68	75	46	48	57
Total errors of identification	1	0	1	8	14	17	2	2	9

More subjects performed correctly (i.e. chose negative instances only) for the simple descriptions than for the disjunctive descriptions (Wilcoxon's $T = 28.5$, $N = 15$, $p < .05$), but there was no difference between the two sorts of disjunctions, and the predicted trend in difficulty across the three types of description was not confirmed (Page's $L = 234.5$, $p > .1$). Additional evidence that subjects had a better grasp of the logic of the task with the simpler descriptions was obtained by comparing subjects' correct and incorrect selections at each of the three levels of difficulty. With the simpler descriptions, subjects selected significantly more negative (i.e. correct) instances than positive or mixed (i.e. incorrect) instances (Wilcoxon's $T = 40$, $N = 19$, $p < .025$), whereas there was no such difference for either type of disjunctive description.

A closer inspection of the pattern of errors revealed that the individual descriptions within each of the three levels of complexity were not of equal difficulty. A test of the difference between two proportions (Hoel, 1971) revealed that the disjunctive description (2a) produced a higher proportion of negative choices than did descriptions (2b) and (2c) ($p < .05$)—the subjects' performance with description (2a) was comparable to that with descriptions (1a) and (1c). In addition, the proportion of negative choices for description (1b) was lower than that for descriptions (1a) and (1c) ($p < .025$).

What is influencing subjects' choices, apart from whether or not a descrip-

tion is disjunctive? A mere disjunction does not necessarily affect insight, since (2a) is easy; nor does a negative description, since (1c) is easy. But, when a disjunction is combined with one or more negatives, insight is reduced—as is shown by performance with descriptions (2b) to (3c). What remains puzzling is the poor performance with description (1b). Since this description has the highest ratio of negative to positive diagrams (7:4), perhaps those subjects who were not certain of their strategy opted to check the smaller set of diagrams, i.e. the positive instances.

We examined the individual protocols to determine at which point the subjects gained insight into the structure of the task and began to choose negative instances exclusively. Twelve subjects first performed correctly with simple descriptions, three with inclusive disjunctions, and four with exclusive disjunctions; the difference between these frequencies was significant ($\chi^2 = 7.73$, $df = 2$, $p < .025$). Although description (2a) generally resulted in correct performance once the structure of the task had become clear, only two subjects gained insight with it. Once subjects had gained insight, it was *never* lost again on encountering the simple descriptions (1a) and (1c).

We plotted the number of times that a description occurred in each serial position over the experiment as a whole, but there was no relation between frequency of occurrence in a particular position, and the level of insight afforded by a description. Each description occurred approximately the same number of times in each third of the experiment. If anything, the simpler descriptions (1a and 1c) occurred more often in the first three trials than elsewhere, and thus the odds were biased *against* the subjects showing insight with these descriptions. Moreover, description (1b) occurred more often towards the end of the experiment. Hence, in general, the level of insight with a particular description does not appear to be related to its serial position in the experiment.

The time that the subjects spent examining the descriptions prior to their first selection was also related to the level of insight. Because there was a fault in the timer, these inspection times were collected for only 17 of the 19 subjects. The mean inspection time for rules that were tested with complete insight was 46 seconds, whereas for rules that were tested with partial (i.e. a mixture of positive and negative choices) or no insight, it was 66 seconds; and the difference was significant ($t = 2.72$, $df = 16$, $p < .01$). There was also a significant difference in the times for the initial selections across the nine descriptions, (see Table 2, $F = 4.59$, $df = 8,144$, $p < .001$). In part, this difference can be attributed to the shorter reading times for the simple descriptions, which also had fewer words but, in addition, the disjunctive descriptions with two negatives (2b and 2c) took significantly longer to read and understand than disjunctive descriptions with only one negative (3a, 3b and

3c), $t = 3.73$, $df = 18$, $p < .005$. The disjunctive descriptions also yielded more errors in identification and, again, descriptions (2b) and (2c) produced more errors than the other disjunctive descriptions.

Discussion

The errors in identification corroborated our assumption that disjunctions are more complex than simple assertions, and confirmed our conjecture that the heavier cognitive load of disjunctions should affect logical insight: subjects reasoned more logically with simple assertions. However, the relation between load and level of insight was not entirely straightforward. Disjunctive descriptions, which necessitated holding two states of affairs in mind whilst ascertaining their negative instances, did not always reduce insight, but only proved troublesome when at least one disjunct was negative. Exclusive disjunctions were no harder than inclusive ones, perhaps because the subjects failed to realise that, in the case of inclusive disjunctions, counterexamples to the more general disjunct were also counterexamples to the more specific one. However, there was some evidence that disjunction alone could cause difficulty: subjects very rarely showed insight *initially* on description (2a). Thus, it seems that the main factor contributing to cognitive load is the need to retain and to evaluate two sorts of description in the case of disjunctions. This difficulty was exacerbated by the well-established problem of understanding negation (see e.g. Clark and Clark, 1977; Wason and Johnson-Laird, 1972).

The phenomenon of apparent loss of insight is particularly surprising since the subjects often verbalised their strategy, but still made errors on subsequent descriptions.

Although we appear to have corroborated Johnson-Laird and Wason's (1970) speculation about cognitive load, another factor that may have affected performance is the ratio of positive to negative instances of a description. When the proportion of negative instances was large, as in the case of description (1b), the subjects again seemed to revert to the incorrect strategy. Moreover, if they had simply selected whichever was the smaller set of diagrams, then they would sometimes have appeared to have had insight into the logic of the task. This factor introduces a potential artefact into the experiment, and we therefore sought to extend its results in a replication in which the number of positive and negative instances was equal for each description.

Experiment 2

Subjects were again asked to test the truth or falsity of descriptions of the contents of an envelope. In order to examine the effects of negation and disjunction, we used four sorts of description: affirmatives and simple negatives, and affirmative and negative disjunctions.

Subjects

Thirty-six subjects participated in the experiment, which lasted 30 to 40 minutes. They were paid £1.50.

Materials and procedure

There were four types of description, with two descriptions of each type:

Simple affirmatives

- 1a. Every dot is connected to at least one other dot.
- 1b. At least one dot is connected to every other dot.

Simple negatives

- 2a. No dot is connected to all the dots.
- 2b. At least one dot is not connected to any other dot.

Affirmative disjunctions

- 3a. Every dot is connected to every other dot
OR at least one dot is connected to every other dot.
- 3b. Every dot is connected to two or more dots
OR all the dots are connected to at least one other dot.









Negative disjunctions

- 4a. No dot is connected to any other dot
OR no dot is connected to all the dots.
- 4b. No dot is connected to two or more dots
OR at least one dot is not connected to any other dot.

Three of the original diagrams were discarded, and for each of the eight rules there were four positive and four negative exemplars in the array of diagrams. Table 3 shows the diagrams and the positive and negative exemplars for each rule.

The procedure was identical to that of Experiment 1, except that the order in which the descriptions were presented was not random. Each type of

Table 3. *The eight diagrams in Experiment 2, and the positive (1) and negative (0) instances of the eight descriptions*

	1	2	3	4	5	6	7	8
								
Descriptions								
1a	0	0	0	0	1	1	1	1
1b	0	0	0	0	1	1	1	1
2a	1	1	1	1	0	0	0	0
2b	1	1	1	1	0	0	0	0
3a	0	0	0	0	1	1	1	1
3b	0	0	0	0	1	1	1	1
4a	1	1	1	1	0	0	0	0
4b	1	1	1	1	0	0	0	0

description appeared once in each half of the presentation, and the order of the four descriptions in each half was based on a Williams' square design: each rule was preceded and followed equally often by every other. A different order was used for the first and second halves of the presentation. There were thus twelve orders in all. In addition, half of the subjects' presentations began with one example of each type of rule, and half with the other example.

Results

Of the 36 subjects tested, 11 showed no insight and 7 showed complete insight into the task from the beginning. The responses from the remaining 18 subjects were classified as in Experiment 1. Table 4 shows the frequency of the three types of response: negative instances only, mixed instances, and positive instances only. Overall, the simple descriptions resulted in a choice of potential counterexamples more often than did the disjunctive descriptions (Wilcoxon's $T = 40.5$, $N = 17$, $p < .05$). The difference between affirmative and negative rules was very small and did not vary as a function of whether the descriptions were simple or disjunctive (Wilcoxon's $T = 56$, $N = 15$, p

> .1). Level of insight was not related to which of each pair of rules was presented: there was no significant difference in the number of negative choices between the two versions for any of the four types of rule.

As in Experiment 1, more subjects (15 out of 18) first showed insight into the task (choosing negative instances only) on simple than on disjunctive descriptions ($\chi^2 = 6.72, p < .01$). Ten subjects reverted to choosing positive or mixed instances after they had performed correctly on at least one description, and these 'losses of insight' were slightly but not significantly more frequent with the disjunctive rules (27% vs. 15%).

The mean inspection times for each type of description are shown in Table 4. An analysis of variance showed only that disjunctive descriptions took longer to read than simple descriptions, $F(1,17) = 36.47, p < .001$. The mean reading time for descriptions yielding complete insight (38 seconds) was significantly lower than the mean for descriptions yielding partial or no insight (45 seconds), $t(17) = 2.08, p < .05$. Since the trials on which insight occurred tended to be towards the end of the experiment, this effect could merely be the result of practice. There were too few errors in identifying instances of descriptions for statistical analysis (see Table 4).

Table 4. *The number of subjects in Experiment 2 selecting negative, mixed and positive instances for each of the eight descriptions; the mean inspection times prior to the first selection; and the total number errors in identifying instances of the descriptions*

	Rules							
	Simple affirm.		Simple negative		Affirm. disjunc.		Negative disjunc.	
	1a	1b	2a	2b	3a	3b	4a	4b
Negative instances	13	11	13	10	9	11	9	9
Mixed instances	3	4	4	5	7	3	5	5
Positive instances	2	3	1	3	2	4	4	4
Mean inspection times (sec.)	31	33	35	26	39	53	40	59
Total errors of identification	0	0	2	0	0	2	0	6

Discussion

The results demonstrate clearly that subjects reasoned more logically when the descriptions were simple rather than complex. The difference cannot be merely a bias towards choosing the smaller set of instances, whether positive or negative, since the proportions were equal for each description. However, we did not replicate our earlier finding that negative disjunctions less often lead to insight than do affirmative ones. The experiment was probably not sensitive enough to detect an effect: the subjects seemed to find the task easier, perhaps because there were fewer diagrams. There were fewer errors of identification than in Experiment 1, more subjects showed complete insight from the outset (19% vs. 10%), and more subjects attained at least partial insight (choosing a mixture of positive and negative exemplars on some trials), even though they never attained complete insight (11% vs. 3%). The four subjects in this latter category, however, *did* find the negative disjunctions particularly difficult: whereas they chose a mixture of positive and negative exemplars in roughly equal proportions for all other rules, they *never* chose any negative exemplars of the negative disjunctions.

General discussion

Although counterexamples play a crucial role in rational thinking, many people do not search for them whether they are formulating a hypothesis, drawing a deductive conclusion, or testing a generalisation (see e.g. Johnson-Laird, 1983; Wason, 1983). There are two potential classes of explanation for this phenomenon: an individual may not grasp the force of counterexamples, or alternatively may experience difficulty in putting the principle into practice. Our results suggest that there are differences in performance from one individual to another.

A few subjects see at once that they must search for counterexamples, and they perform this task without error. Other subjects, at the opposite end of the spectrum, fail to appreciate the relevance of counterexamples, and persist in selecting positive instances throughout the experiment. They never select anything that could falsify a description, and thus they never learn the folly of seeking only confirmation. Many subjects, however, begin to make rational selections at some point during the experiment. These subjects are obviously the most interesting, because only they can illuminate the process of gaining insight into the importance of counterexamples and the causes of difficulty in putting the principle into practice. It is reasonable to suppose that some, if not all, of them have a partial grasp of the principle in daily

life. This grasp is not sufficient to ensure that they start the experiment with complete insight, but it does ensure that after some experience with the task they realise the relevance of counterexamples.

The most striking of our findings is that subjects tend to gain insight with a simple description, and if they subsequently lose it, they are more likely to do so with a disjunctive description. The effects of disjunction may be twofold. First, such a description may increase the load on working memory in the same way as having to retain a six-digit number whilst reasoning (see Baddeley and Hitch, 1974; Hitch and Baddeley, 1976). Second, a disjunctive description is evidently harder to understand—in both experiments, subjects spent longer inspecting disjunctions before they made their first selection, and they made more errors in identifying instances of the disjunctive descriptions in Experiment 1 (there were hardly any such errors in Experiment 2, perhaps because the task was made easier by the smaller array of diagrams). Hence, the root of the problem is likely to be in determining what counts as a counterexample of a description. A parsimonious explanation of the difficulty of disjunctions is suggested by the theory of ‘mental models’, which has been successfully applied to spatial, relational, and syllogistic reasoning (see e.g. Johnson-Laird, 1983): it is harder to keep two models in mind than one. This conjecture received informal support from the comments of some of our subjects. They remarked on their attempts to transform the rules into ‘characteristic diagrams’, and on the need to produce two such ‘diagrams’ for disjunctions. Such observations give us a useful clue to why disjunctive descriptions are so detrimental to reasoning (see e.g. Bruner, Goodnow and Austin, 1956; Newstead and Griggs, 1983; Wason, 1977): they require the subject to keep in mind two different ‘prototypical’ diagrams and then to find counterexamples to both of them, whereas a simple description requires the reasoner only to retain one representative model and then find a counterexample to it. When our subjects’ working memory is employed in the retention of two different models, they have less processing capacity available to determine a rational search strategy.

A factor that is claimed to affect performance in testing generalisations is previous experience: one either directly recalls the class of counterexamples (Manktelow and Evans, 1979) or else an analogous memory somehow triggers a process leading to their selection (Griggs, 1983). Direct memory for counterexamples fails, of course, to explain performance in our experiments. Subjects undoubtedly had never before encountered, say, the generalisation: ‘At least one dot is connected to every other dot.’ Had they encountered an analogous generalisation? Perhaps. But the weakness of this hypothesis is precisely that it provides us with no way of answering the question decisively, since it does not specify what counts as an analogous generalisation. Because

any effect of content per se must ultimately be a function of memory, the most that we can safely conclude is that previous experience may sometimes make it easier to select counterexamples. This hypothesis, however, is entirely compatible with our view that the critical factor is the ease of determining what counts as a counterexample to a generalisation. Anything that makes this task easier is likely to improve performance. This thesis has received independent corroboration in a recent study carried out by Wason and Green (1984). They showed that subjects search for counterexamples to coherent generalisations, such as descriptions of figures and their grounds:

Whenever they are triangles they are on black cards

to a greater degree than they search for counterexamples to disparate descriptions, such as:

Whenever there are triangles below the line, there is black above the line

Rationality depends on a search for counterexamples. If, say, you hold the prejudice that women are bad drivers, and your curiosity about gender is only provoked by cases of bad driving, then you will never be shaken from your bias: if a bad driver turns out to be a woman, your prejudice is confirmed; if a bad driver turns out to be a man, your prejudice is not disconfirmed since you don't believe that only women are bad drivers. Unless you somehow are able to grasp the potential relevance of *good* drivers to your belief, then the danger is that you will never be disabused of it, and will never understand the force of counterexamples. The moral of our research is that the easier it is to determine what would count as counterexamples to a generalisation, the more likely reasoners are to appreciate the need to search for them, and to maintain that insight.

References

- Baddeley, A.D. and Hitch, G.J. (1974) Working memory. In G. Bower (ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 8. New York, Academic Press.
- Bruner, J.S., Goodnow, J.J. and Austin, G.A. (1956) *A Study of Thinking*. New York, Wiley.
- Clark, H.H. and Clark, E. (1977) *Psychology and Language*. New York, Harcourt.
- Evans, J.St.B.T. (1982) *The Psychology of Deductive Reasoning*. London, Routledge and Kegan Paul.
- Griggs, R.A. (1983) The role of problem content in the selection task and THOG problem. In J.St.B.T. Evans (ed.) *Thinking and Reasoning: Psychological Approaches*. London, Routledge and Kegan Paul.
- Griggs, R.A. and Cox, J.R. (1982) The elusive thematic-materials effect in Wason's selection task. *Brit. J. Psychol.*, 73, 407-420.
- Hitch, G.J. and Baddeley, A.D. (1976) Verbal reasoning and working memory. *Q. J. exp. Psychol.*, 28, 603-621.

- Hoel, P.G. (1971) *Elementary Statistics*. (3rd edn.) London, Wiley.
- Johnson-Laird, P.N. (1983) *Mental Models*. Cambridge, Cambridge University Press.
- Johnson-Laird, P.N. and Steedman, M. (1978) The psychology of syllogisms. *Cog. Psychol.*, 10, 64-98.
- Johnson-Laird, P.N. and Wason, P.C. (1970) Insight into a logical relation. *Q. J. exp. Psychol.*, 22, 49-61.
- Manktelow, K.I. and Evans, J.St.B.T. (1979) Facilitation of reasoning by realism: effect or non-effect? *Brit. J. Psychol.*, 71, 227-231.
- Newstead, S.E. and Griggs, R.A. (1983) The language and thought of disjunction. In J.St.B.T. Evans (ed.) *Thinking and Reasoning: Psychological Approaches*. London, Routledge and Kegan Paul.
- Wason, P.C. (1960) On the failure to eliminate hypotheses in a conceptual task. *Q. J. exp. Psychol.*, 12, 129-140.
- Wason, P.C. (1977) Self-contradictions. In P.N. Johnson-Laird and P.C. Wason (eds.) *Thinking: Readings in Cognitive Science*. Cambridge, Cambridge University Press.
- Wason, P.C. (1983) Realism and rationality in the selection task. In J.St.B.T. Evans (ed.) *Thinking and Reasoning: Psychological Approaches*. London, Routledge and Kegan Paul.
- Wason, P.C. and Green, D.W. (1984) Reasoning and mental representation. *Q. J. exp. Psychol.*, 36A, 597-610.
- Wason, P.C. and Johnson-Laird, P.N. (1972) *Psychology of Reasoning: Structure and Content*. London, Batsford.

Résumé

Deux expériences ont été faites pour étudier les raisons des échecs dans la sélection de contre-exemples permettant de vérifier des généralisations. Les sujets doivent déterminer si la description du contenu d'une enveloppe (une série de diagrammes) est correcte ou fausse. Etant donné qu'un diagramme correspondant à la description peut se trouver soit à l'intérieur soit à l'extérieur de l'enveloppe sans affecter le statut de la description, la stratégie rationnelle est de choisir les diagrammes qui ne correspondent pas à la description puisqu'ils peuvent en principe falsifier celle-ci. Les sujets sélectionnent les diagrammes à partir de deux rangées situées en face d'eux et l'expérimentateur identifie chaque choix comme "dans ou hors de l'enveloppe". Il apparaît avec la première expérience que les descriptions disjonctives posent des problèmes: les sujets en tirent peu d'informations et souvent perdent leur intuition première. La présence d'une négation dans la description amplifie ces effets. Dans une seconde expérience on a examiné systématiquement l'effet des disjonctions et des négations en contrebalançant le nombre des exemples positifs et négatifs dans chaque description. On retrouve les effets de la disjonction mais pas ceux de la négation. Les patterns des erreurs et les temps de latence suggèrent une interprétation générale de ces échecs: les sujets ne voient pas immédiatement la nécessité de rechercher des contre-exemples et perdent cette intuition ce qui rend difficile de déterminer ensuite les contre-exemples possibles.