

REASONING WITH AMBIGUOUS SENTENCES

By P. N. JOHNSON-LAIRD

Department of Psychology, University College, University of London

Eight inferences with quantified sentences were judged as valid or invalid. Each inference involved an active and its correlative passive sentence, e.g. 'Some medicine cures every disease. Therefore: Every disease is cured by some medicine.' There were four logically distinct pairs of sentences, and each pair was presented twice: once as an inference from active to passive, and once as the converse inference from passive to active. As predicted, what was crucial in evaluating these inferences was not voice but the position of 'some'. An inference from a premise with 'some' in the grammatical subject to a conclusion with 'some' in the grammatical object tended to be judged as valid, whereas an inference in the converse direction tended to be judged as invalid. The results of an earlier study of the ambiguity of the sentences provided a good estimate of performance in the present inferential task.

There is a long-standing tradition that the layman is a poor reasoner, prey to a large number of distractions—from the 'atmosphere' of words (e.g. Woodworth & Sells, 1935) to the bias of prejudice or emotion (e.g. Kaufmann & Goldstein, 1967). From time to time attempts are made to rehabilitate the layman's reputation (e.g. Henle, 1962), but with no great success. An alternative tradition, however, suggests that natural language is too ambiguous a vehicle for consistent reasoning (e.g. Russell, 1957). It is a corollary to this view which is the basis of the present investigation: if an inference is couched in ambiguous sentences, then the degree of disagreement about its validity should be accounted for by the degree of ambiguity.

One of the drawbacks of earlier work on reasoning is that no attempts were made to discover just what the material was understood to mean. Indeed, subjects were sometimes *instructed* how they were to interpret sentences, e.g. a sentence of the form 'some plants are green' was *not* to be interpreted to imply that 'some plants are not green'. Such instructions may run counter to the spontaneous interpretation of the sentences, and are liable to cause difficulty. A preferable strategy is to allow subjects to interpret the sentences in their own way, but to obtain some independent evidence of just what interpretations were likely to have been made. This strategy was indispensable in the present study, where the sentences were both ambiguous and logically complex.

Certain active and passive sentences are ambiguous in an interesting way. Consider, for example, these two sentences: (1) 'Every child loves some toy.' (2) 'Some toy is loved by every child.' At first sight, sentence (1) seems to mean that every child loves some toy or other, whereas sentence (2) seems to mean that every child loves the same *particular* toy. On further reflexion, however, one realizes that both sentences are ambiguous and may receive either interpretation (Katz & Postal, 1964, p. 72). Nevertheless, the sentences do differ in their most likely interpretations, and one consequence of this difference is that the inference: (1) 'Every child loves some toy. Therefore: Some toy is loved by every child' should tend to be judged invalid, whereas the converse inference: (2) 'Some toy is loved by every child. Therefore: Every child loves some toy' should tend to be judged valid, even though both inferences involve identically ambiguous material. These predictions follow from the more likely inter-

pretations of the sentences. And it has been shown in an earlier study (Johnson-Laird, 1969) that the 'some in particular' interpretation tends to be given when 'some' occurs at the beginning of the sentence in the grammatical subject, whereas the 'some or other' interpretation tends to be given when 'some' occurs at the end of the sentence in the grammatical object. This is so, regardless of the voice of the sentence, and it would seem that the actual position in the sentence is more important than grammatical status (Johnson-Laird, 1968). Given these interpretations of 'some' exactly the same predictions apply to inferences involving negative sentences. Hence the predictions may be concisely stated: an inference from a premise with 'some' in the subject to a conclusion with 'some' in the object should be judged as valid, and an inference in the converse direction should be judged as invalid. It was this prediction which was tested, by examining eight logically distinct inferences.

Sentences like (1) and (2) express what logicians call 'quantified relations', and since they contain two quantifiers (e.g. 'some', 'every') they denote a 'doubly quantified' relation (cf. Suppes, 1957). There are 12 such relations, any one of which may be expressed in English in a variety of ways. Eight were investigated in the earlier study because a fairly conventional, though ambiguous, representation of them is possible, using four pairs of active and passive sentences. The same material was used in the present study, but in the form of inferences—from active to passive and from passive to active—in order to discover to what extent performance between the two tasks was compatible.

METHOD

Design. Each subject, acting as his own control, had to judge the validity of eight inferences, which are given below in a standard lexical format. In practice, however, each inference was presented to the subject with a different lexical content. The order of presentation of the different types of inference was counterbalanced using 24 orders derived from three 8×8 Williams squares, with a different order of presentation of the eight lexical contents being assigned to each square. This provided an adequate control for any interaction between form and content.

In order to provide a greater variety of material for the subject to evaluate, and thus inhibit the development of any stereotyped approach to the task, four 'filler' items of a different form were also introduced. They were the same for all subjects and were presented in the same order, on the first trial and thereafter, following every second test inference.

Materials. Each pair of sentences was used to form two inferences, yielding the following eight types of inference, stated with their predicted evaluations in parentheses:

- (1) Every man knows some woman.
Therefore: Some woman is known by every man. (Invalid.)
- (2) Some woman is known by every man.
Therefore: Every man knows some woman. (Valid.)
- (3) Every woman is known by some man.
Therefore: Some man knows every woman. (Invalid.)
- (4) Some man knows every woman.
Therefore: Every woman is known by some man. (Valid.)
- (5) Any man does not know some woman.
Therefore: Some woman is not known by any man. (Invalid.)
- (6) Some woman is not known by any man.
Therefore: Any man does not know some woman. (Valid.)
- (7) Any woman is not known by some man.
Therefore: Some man does not know any woman. (Invalid.)
- (8) Some man does not know any woman.
Therefore: Any woman is not known by some man. (Valid.)

The lexical material for these inferences was chosen so as to vary types of noun (e.g. \pm human, \pm abstract, etc.) and the 'reversibility' of the sentences, i.e. whether the two nouns can be interchanged without producing a semantic anomaly. The eight lexical contents were as follows:

- (Every) man knows (some) woman.
- (Every) car overtakes (some) bus.
- (Every) child likes (some) toy.
- (Every) job requires (some) skill.
- (Every) brother hates (some) sister.
- (Every) liquid dissolves (some) solid.
- (Every) medicine cures (some) disease.
- (Every) animal possesses (some) instinct.

The 'filler' items involved sentences about which there was no interpretative data, but which were of independent interest:

- (1) Every dish needs no sauce.
Therefore: No sauce is needed by every dish.
- (2) No philosopher understands every theorem.
Therefore: Every theorem is understood by no philosopher.
- (3) No girl is teased by every boy.
Therefore: Every boy teases no girl.
- (4) Every crime is punished by no law.
Therefore: No law punishes every crime.

The test materials involved 64 different sentences (eight logical forms \times eight lexical contents), and the 'filler' materials involved eight different sentences. Each sentence was typed separately on a plain 5 \times 3 in. card.

Subjects. The subjects were 24 undergraduates at University College London. They were native speakers of English, unfamiliar with the quantificational calculus, and had not taken part in any previous investigation of this type.

Procedure. The subjects were successively allocated to a particular combination of logical and lexical material. They were tested individually.

The main feature of the procedure was that premise and conclusion were always separately presented, and at no time did a subject have both in front of him. This was designed to minimize a purely 'syntactic' approach to the task, which would be more likely to occur if subjects noted the relation between premise and conclusion.

The following instructions were read aloud by the experimenter: 'This experiment concerns reasoning—it's not about *your* ability to reason. I merely want to find out which of certain inferences are generally considered to be valid. A valid inference can be made from false premises. For example, if I assert that Hitler is alive, then you may validly infer that Hitler is not dead. This is obviously a trivial inference, but nevertheless one sees that there is a logical connexion between the two statements, which is there even though premise and conclusion are false. The inferences that I'm going to ask you to evaluate will also be from one sentence to another; and I want you to forget all about whether the sentences are true or false and just consider whether one follows logically from the other. O.K.? I'll show you the premise first, and you may study it for as long as you like, until you're satisfied that you grasp its implications. Then I shall remove it and place in front of you the conclusion. It's your task to decide whether the conclusion follows logically from the premise. I'm not timing you, so take your time and make sure that your judgement reflects your considered opinion.'

It was then explained that the subject could not refer back to the premise once it had been removed, nor could he ask any questions about the sentences. Judgements were to be indicated by saying 'valid' or 'not valid'.

RESULTS

Test inferences

Each subject was given a score equal to the number of judgements he made in accordance with the prediction. The frequencies with which each possible score occurred are shown in Fig. 1. The average number of such judgements was 5.2 (out of a maximum of 8) and this bias in favour of the prediction was significant on a Wil-

coxon test ($P < 0.005$, one-tailed). Fig. 2 shows the number of subjects making the predicted judgement for each type of inference.

In order to test whether the distribution of judgements could be accounted for by the practical ambiguity of the sentences, the results were compared with those from the earlier interpretative study. The evaluation of two related inferences, one from

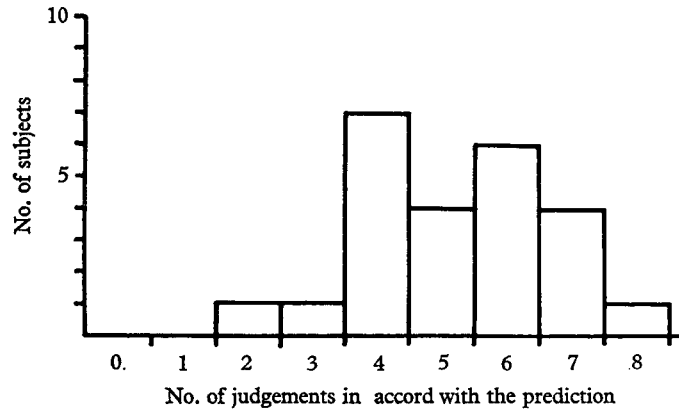


Fig. 1. The distribution of scores (maximum possible score = 8).

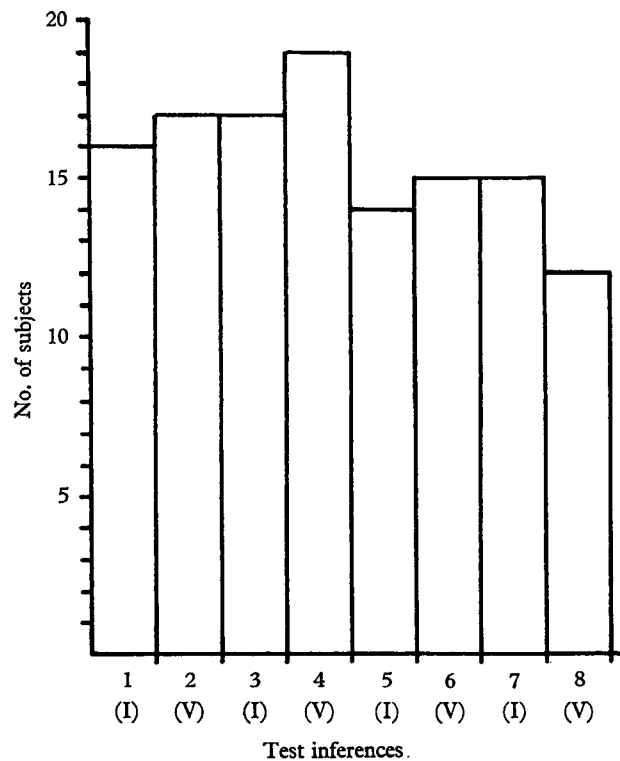


Fig. 2. The number of subjects making the predicted judgement for each type of inference. V, predicted judgement of 'valid'; I, predicted judgement of 'invalid'.

active to passive and the other from passive to active, must result in one of four possible outcomes:

- (a) The sentences are judged to follow from one another, i.e. both inferences are judged as valid.
- (b) The sentence with 'some' in the object is judged to follow from the sentence with 'some' in the subject, but not conversely.
- (c) The sentence with 'some' in the subject is judged to follow from the sentence with 'some' in the object, but not conversely.
- (d) Neither sentence is judged to follow from the other, i.e. both inferences are judged as invalid.

The number of subjects in each of these outcomes, for each related pair of inferences, is shown in Table 1, with the comparable figures from the interpretative study. The

Table 1. *The number of subjects in each outcome contingency for each related pair of inferences, and the comparable results from the earlier interpretative experiment*

Inferences...	(i) 1 and 2				(ii) 3 and 4				(iii) 5 and 6				(iv) 7 and 8			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
Interpretation	2	16	2	4	1	12	2	9	1	8	3	12	4	13	3	4
Inference	5	12	3	4	5	14	2	3	5	10	5	4	7	5	2	10

homogeneity of the results was estimated heuristically by computing χ^2 for each matrix. The following values of χ^2 (3 d.f.) were obtained: (i) 2.06 ($P > 0.5$); (ii) 5.81 ($P > 0.1$); (iii) 7.38 ($P > 0.05$); and (iv) 7.12 ($P > 0.05$); hence there was no markedly significant departure from homogeneity.

Introspective reports

The majority of those subjects who were able to say how they had performed the task implied that they had relied upon verbal methods, e.g. 'I worked out the meaning of the second sentence by comparing it with the first', 'I compared whether the conclusion was saying the same thing as the premise'. But four subjects claimed to have used visual images in judging the inferences: 'I tried to get a picture—it's more concrete than verbal meaning', 'With the sentence about some woman, I picture the little woman in the street, who thousands of little men loved (*sic*), ... then I compared the two pictures'.

Two subjects said spontaneously that they had soon realized that premise and conclusion were grammatically related. Two other subjects showed partial insight: 'When I got the second sentence, I tried to recall the first; in most cases it was a matter of reversing the second statement and then comparing it with the first', 'Halfway through, I felt I could guess what the conclusion would be'. The majority of subjects, however, did not seem to be aware of the relation; and the subjects with insight had no bias towards judging the inferences as valid. On the contrary, their judgements were biased towards the 'invalid' category (21:11).

A number of things were blamed for causing difficulty in the task. Most frequently subjects mentioned the word 'some', 'any' was sometimes coupled with it; but only three subjects explicitly mentioned negatives. One subject said that to minimize difficulty he had converted sentences of the form 'Any child does not love some toy'

to those of the form 'No child loves every toy'—the probable interpretations of both these sentences are the same. Subjects also mentioned ambiguity as a cause of difficulty, with varying degrees of specificity: 'Certain sentences were ambiguous', 'It was especially difficult if both sentences were ambiguous', 'Sentences with "some" and "any" were ambiguous', "'Some woman is known ...'" means a particular woman unlike "Everyone knows some woman"', "'Some" seemed to mean either no one in particular or a particular one'. There were a few occasions when the lexical content of the sentences seemed to have caused difficulty, either because it referred to something implausible or because it engaged subjects' emotions.

DISCUSSION

This experiment has shown that subjects are capable of exercising logical acumen in judging inferences with 'doubly quantified' sentences. Unlike *making* an inference, such a task involving merely the evaluation of inferences does not necessarily elicit any process of reasoning. The subjects might guess their answers or, at a more sophisticated level, they might base them upon whether they thought the conclusions were true. In fact, subjects showed a consistent bias towards the predicted outcomes, and this can only be explained by some form of reasoning. Since the judgements were to a considerable extent governed by the predicted interpretations of the sentences, they support the underlying hypothesis about word order: 'some' in the grammatical subject tends to be interpreted as 'some in particular', and in the grammatical object it tends to be interpreted as 'some or other'.

When the results are compared with those from the earlier study (Johnson-Laird, 1969) it is clear that the interpretative data are an even more accurate guide to performance than the theoretical predictions. Unaided intuition cannot provide estimates of the probabilities of the various interpretations, and the empirical estimates are certainly superior to those obtained by assuming some fixed probability over all the sentences.

There was a tendency, evident in all pairs of inferences (see outcome (a) in Table 1), for the correlated sentences to be judged to follow from one another more often than would be expected from the interpretative data. This suggests that the two sentences were given synonymous interpretations, and that subjects were biased towards giving the conclusion the same interpretation as the premise. In other words, there was a *semantic atmosphere effect*: the interpretation of the premise influenced the interpretation of the conclusion. However, there is a possible alternative, though related, explanation. Having judged one inference as valid, a subject might be biased towards judging the other related inference, occurring later in the series, as also valid. But, if this were so, a similar effect should occur when the first inference was judged *invalid*; hence, there should also be a greater tendency for neither sentence to be judged to follow from the other than would be expected from the interpretative data. Table 1, outcome (d), shows that there was in fact no such tendency, and this implies that the atmosphere effect is primarily confined to processes within the inference. However, it is by no means clear to what extent it is due to some incidental difference between the interpretative and inferential tasks. Future research must attempt to establish the phenomenon unequivocally, and one immediate goal, prior to the examination of

more complicated inferences, is to determine the effects of ambiguity when premise and conclusion are *not* grammatically related, e.g. 'Some man does not know any woman. Therefore: No woman is known by every man'. The semantic atmosphere effect may diminish if it depends partially upon a grammatical relation. One final piece of evidence that suggests that context is crucial is that more subjects noticed the ambiguity of the sentences in the inferential task than in the interpretative task. Yet ambiguity was far more important in the interpretative task. It seems, therefore, that it was more noticeable when a context and contrast was provided by the other sentence in the inference.

It is evident from Table 1 that there was some deterioration in the homogeneity of the results from the two experiments, when the materials involved negatives. One possible factor is the extra difficulty of negative sentences (cf. Wason, 1961). Another possible factor is that, unlike the other sentences, those of the form 'Any ... not ... some ...' are, in practice, six ways ambiguous (cf. Johnson-Laird, 1969). Either or both of these factors could lead to a degradation in logical performance; and a plausible alternative strategy, which is compatible with the evidence, is for the subjects to base their evaluation of the inference upon whether the premise and the conclusion are synonymous.

The original motivation of this study was to rehabilitate the layman's reputation as a rational thinker and to show that, in performance, natural language is insufficiently stable for consistent reasoning. Such an all-embracing theory is hardly to be demonstrated empirically, and it was necessary to test the more limited proposition that the ambiguity of the material would account for the judgements of the inferences. This proposition has received empirical support, and even when performance departs from it—in the guise of a semantic atmosphere effect—this is only made possible by the ambiguity of language. It is often remarked by psychologists that there is something unreal about investigations of sentences isolated from their contexts. The present study lends support to this view in so far as it indicates that the way one sentence is interpreted influences the way another is interpreted.

The author is extremely grateful to Dr P. C. Wason for his ideas and encouragement during all stages of the work, and to Dr A. R. Jonckheere for advice stemming from a number of stimulating discussions.

REFERENCES

- HENLE, M. (1962). On the relation between logic and thinking. *Psychol. Rev.* **69**, 366–378.
- JOHNSON-LAIRD, P. N. (1968). The choice of the passive voice in a communicative task. *Br. J. Psychol.* **59**, 7–15.
- JOHNSON-LAIRD, P. N. (1969). On understanding logically complex sentences. *Q. J. exp. Psychol.* **21** (in the Press).
- KATZ, J. J. & POSTAL, P. M. (1964). *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: M.I.T.
- KAUFMANN, H. & GOLDSTEIN, S. (1967). The effects of emotional value of conclusions upon distortions in syllogistic reasoning. *Psychon. Sci.* **7**, 367–368.
- RUSSELL, B. A. W. (1957). Mr Strawson on referring. *Mind* **66**, 385–389.
- SUPPES, P. (1957). *Introduction to Logic*. London: Van Nostrand.
- WASON, P. C. (1961). Response to affirmative and negative binary statements. *Br. J. Psychol.* **52**, 133–142.
- WOODWORTH, R. S. & SELLS, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *J. exp. Psychol.* **18**, 451–460.

(Manuscript received 27 April 1968)