IN REAL LIFE NEGATIVES ARE FALSE

P. C. WASON

Psycholinguistics Research Unit
University College London

Summary

This paper in concerned with the psychological problem of negation rather than with any philosophical side-effects which may be discerned in it. The research falls naturally into four phases, but the essence of the problem can be put in a few words.

Negatives do not usually cause difficulty in daily life, but in laboratory tasks the subjects tend to adopt computer-like operations in order to avoid considering them directly. This apparent paradox may be resolved by an interpretation of the fact that, in verification tasks, true affirmatives are easier than false affirmatives, but false negatives easier than true negatives. Formal models of negation explain the coding processes responsible for this interaction, but ignore the fact that the "false negative", rather than the "true negative", approximates to linguistic usage. The "false negative" is analogous to the negative which corrects a preconception rather than one which maintains a truth. It is suggested that the serial operations, which the subjects adopt in verification tasks, are an attempt to recover the preconception when it is not explicit. It follows that such tasks are misleading because they evaluate the response to sentences in relation to facts rather than preconceptions. Some recent experiments which seek to correct this bias are discussed.

IN REAL LIFE NEGATIVES ARE FALSE

In one sense the title of this paper is plainly wrong, but in another sense I shall argue it is right. It crucially hinges on

the function which negation generally performs in language, and the standards against which its understanding should be evaluated. I should not count facts among these standards, but that is to anticipate my story by a long way.

The research, which ranges from 1959 to 1971, is coloured by my own interpretation of results. This may seem unconventional in a scientific paper, but it must be remembered that the faceless objectivity of much scientific prose often conceals an unconscious selection of facts to fit the author's bent. So what I shall offer is a rough sketch of a changing background, which may have some intrinsic interest because of the span of time it covers, and a more detailed reinterpretation of a psychological problem. This problem may have some relevance for epistemology and logic, but philosophers can best judge that from a straight psychological account of our efforts, unaided by amateur speculation. It will be apparent that what may seem so obvious now was hidden from us a dozen years ago, but this fact will be surprising only to those unacquainted with scientific research. The work falls into four conceptual phases.

(1) *Exploration*

In the 1950's psychological research on language was dominated by information theory (e.g. Miller, 1951). Cognitive psychologists in those days did not talk about linguistics, let alone grammar. "Psycholinguistics" was an unknown word to us, and the pioneering investigations of Frieda Goldman Eisler into hesitations in spontaneous speech were not yet widely known (Goldman Eisler, 1968). However, in 1953 I was an industrial psychologist, and knew very little about systematic research on language. In collaboration, first with A. R. Jonckheere, and later with Sheila Jones, I had developed and tested an algorithmic technique for clarifying incomprehensible leaflets concerned with social security. Among other advantages, it totally eliminated negative sentences: it is logically equivalent to affirm a negative and to deny its correlative affirma-

tive (e.g. Wason, 1962; Jones, 1968). Was there, I wondered, a psychological equivalence ? However, my notebook reveals only the vaguest apprehension of a problem at this time: "Is it easier to discriminate negative information which is fulfilled by the facts, or positive information which is not fulfilled by the facts ?" The recent publication of *A study of thinking* (Bruner, Goodnow and Austin, 1956) had already stimulated my interest in the difficulty of using negative instances in concept attainment, but apart from this line of research, I knew only that Freud and Piaget had made some observations about negation, but these seemed hardly relevant to the kind of experiment which I was in the process of conceiving. Intuition had to take the place of theory.

In the first experiment (Wason, 1959), the subjects had to construct affirmative and negative sentences, in accordance with an instruction to make them true or false, in relation to a changing display of coloured discs. The dependent variables were the times and errors made in doing this. In effect, four basic types of sentence were distinguished:

(1) True affirmatives (TA),	e.g. "There is red"	(red present)
(2) False affirmatives (FA),	e.g. "There is green"	(red present)
(3) True negatives (TN),	e.g. "There is not green"	(red present)
(4) False negatives (FN),	e.g. "There is not red"	(red present)

Thus, the task had three components: an instruction, a sentence and a state of affairs. In fact, it was rather more complex than that, but for purposes of exposition we can assume that the subject had to select the appropriate predicate (i.e. colour word) from a pair contained within each sentence. This procedure makes apparent the logical equivalence between sentence types (1) and (4) on the one hand, and types (2) and (3) on the other hand. Six examples of each of the four types were presented. The errors were relatively few (except for FN when more than half the subjects made an error on its first presentation), but they correlated with the response times. These were in the same order at each of the six presentations, and they were reliably distinguished, their overall mean values

being: TA = 8.99 sec., FA = 11.09 sec., TN = 12.58 sec., FN = 15.17 sec.

I did not really believe these results, and moreover my task appeared to have a defect. The information conveyed by two of the conditions (FA and TN) was less specific than that conveyed by the other two (TA and FN). The influence of information theory (and my colleagues) persuaded me that the amount of information conveyed by each condition should have been equated in order to allow any valid comparisons. It will be remembered that this was exactly the criticism which Hovland and Weiss (1953) had made of Smoke's (1933) experiments on the effect of negative instances in concept formation. Accordingly, a binary task was devised to correct this defect. (I know now that it was no defect at all, and that a binary task has the more serious fault that the subject can "convert" a negative sentence into an affirmative one by mentally transposing the predicate).

The experiment (Wason, 1961) used sentences which referred to the "evenness", or "oddness", of whole numbers. Two types of task were carried out: (1) a *construction task*, formally analogous to that used in the previous experiment, and (2) a *verification task*, which consisted in determining whether a given sentence about a number was true or false. In the construction task the subjects had to name a number to make an incomplete affirmative, or negative, sentence true or false, in accordance with instruction. In the verification task they had to respond with the words "true", or "false", to each sentence, e.g.

(1) TA: "24 is an even number"
(2) FA: "39 is an even number"
(3) TN: "57 is not an even number"
(4) FN: "92 is not an even number"

The results in the construction task exactly corroborated those in the 1959 study. But in the verification task only the negaties were reliably slower than affirmatives; their truth value apparently made no difference. We know now that this result

was an artefact, due either to insensitive timing, or to a mixture of two different strategies on the part of the subjects. In fact, in verification tasks there is a highly corroborated interaction between the four conditions: true affirmatives are faster than false affirmatives, but false negatives are faster than true negatives, i.e. TA<FA<FN<TN (Wason and Jones, 1963; McMahon, 1963; Gough, 1965; Slobin, 1966). This interaction is the heart of the negation problem, and it remained a mystery for some time. The startling qualitative observations held my interest.

We shall disregard the construction task, which has assumed less theoretical importance, and concentrate on the subjects' introspections in the verification task. They subsequently reported that, when a negative sentence was presented, they either mentally deleted the word "not", then verified the resulting affirmative, and finally gave the opposite answer as the response, or they read "not even" as "odd" and "not odd" as "even", and then directly verified the resulting affirmative. These tortuous procedures seemed all the more remarkable because the subjects obviously knew the *meaning* of the negative sentences. They evidently did not try to understand them; they simply refused to accept them. Like a computer programmed with an extra operation for negation, they turned them into affirmatives.

I suggested that this avoidance of negation might be due to (1) an under-estimation of negative information in early conceptual learning, and (2) to the emotional connotations of the word "not", generalised from prohibitives. There was little that we could do about (1), but we tackled (2) by trying to separate the denotative and connotative effects of negatives (Wason and Jones, 1963). The experiment misfired in its aims (the fault being Wason's and not Jones'). We thought we could train the subjects to construe one neutral sign as functioning to assert a proposition, and another as functioning to deny it. With a little more wisdom I might have anticipated the result: the subjects either interpreted the signs linguistically (they were then hardly distinguishable from their English counterparts), or they incorporated them into algorithms which had

no connection with the functions of assertion and denial. An instructive failure. The "emotionality hypothesis" has attracted little attention, perhaps because it is difficult to test, but some very recent research by Ben Reich (1970) has lent it some credence. More compelling evidence may eventually be found in clinical research.

But whatever the reason for the avoidance of negation in an innocuous task, the serial step-by-step processing of a negative sentence, and the affirmation-negation — truth-falsity interaction merited serious attention because it was totally unexpected. Goldman Eisler and Cohen (1970) have recently observed the infrequency of explicit negatives in spontaneous speech, and this could be related to the relative slowness with which they were grasped in my experiments; it could not account for the procedures invoked to deal with them. After completing these early studies, however, I was blocked by the problem, and turned my attention elsewhere. But in 1962 they received a much needed theoretical boost. It was the first time that I (and probably many others) had thought of syntax in relation to psychological research. My vague knowledge of it was derived from tutorials in English philology, and it had seemed to me singularly dull.

(2) *Syntax.* In the early 1960's George Miller and his associates at Harvard were investigating whether an early version of Chomsky's transformational grammar (1957) had a psychological reality. The linguistic theory represented what seemed at that time to be the most economical axiomitization of a grammar. It postulated that certain "optional transformations", e.g. passives, negatives and interrogatives, may be derived from "kernel" structures (simple, active, affirmative, declaratives) which are assumed to be more primitive. Miller (1962) carried out a series of novel experiments to see whether such transformations could be detected in the subjects' performance. For example, given the sentence, "John liked the small boy", it was confirmed that the time taken to identify its negative transform, "John didn't like the small boy", and its passive transform, "The small boy was liked by John", summated for

the time taken to identify the passive-negative transform, "The small boy wasn't liked by John". Similar results were obtained when the subject started with a transformed sentence and was instructed to identify the kernel, and it was also shown that other grammatical features, e.g. verb form, had no such summative effects (Miller and McKean, 1964). Thus grammatical transformations seemed to be both serial and additive. At the syntactical level, empirical psychology had demonstrated that grammatical rules were reflected in operations which could be independently measured.

Miller appreciated that the results of my experiments, which had also provided evidence for serial operations, might have been due to the effects of such transformations. Accordingly, Lee McMahon (1963) adapted the verification task for these ends. The subjects had to judge whether sentences were true or false in relation to a picture, or a known fact, e.g. "5 is followed by 13" (in the number series). The additive relation between the transformations was generally confirmed, but some unpredicted results were obtained. In the Miller and McKean experiments it was found that the syntactically more complex passive transform took about half a second longer than the negative one. McMahon found the opposite: the difference between actives and passives took about 0.1 sec., but the difference between affirmatives and negatives took about 0.4 sec. Indeed, McMahon concluded that: "Through all of the experiments negativity, as such, regardless of grammatical complexity, had effects on behaviour". It seemed clear that these effects must be attributable to semantic rather than syntactic factors — a conclusion which was also acknowledged by Miller (Miller and McNeill, 1969). McMahon's sentences had demanded more than matching on a purely verbal basis.

His results also provided the first confirmation of the affirmation-negation — truth-falsity interaction which we had already obtained (Wason and Jones, 1963). Neither he, nor I, could explain it. A subsequent investigation by Dan Slobin (1963, 1966), in which both children and adults were tested, yielded new results with respect to passives, as well as the incidental

finding that several of the youngest children refused to accept a negative sentence as being true at all. Since he also obtained the interaction I felt compelled to formulate a model to account for it. It assumed that a negative contributed more difficulty than a mismatch between a sentence and its referent. This was represented numerically by assigning a weight of unity to a mismatch, and a weight of two to a negative, with an affirmative and a match weighted as zero. The model (Slobin, 1963) predicts the order of difficulty of the four types of sentence which had now been obtained by all of us in verification tasks:

Sentence type	Negativity	+	Mismatch	=	Difficulty
TA	0	+	0	=	0
FA	0	+	1	=	1
FN	2	+	0	=	2
TN	2	+	1	=	3

The model is crude and *ad hoc*. Indeed, Slobin had some difficulty in making it fit his data from passives. Its chief value at that time was the heuristic one of enabling us to see that the "false negative" in these tasks is not (as its name might suggest) a "double negative". Subsequently I formulated a more intuitive representation:

TA = "a fact"
FA = "a falsehood"
FN = "denial of a fact"
TN = "denial of a falsehood"

Taking all these results, which had been obtained at Harvard's Center for Cognitive Studies in the early 1960's, it could be inferred that both syntactic and semantic factors affected our understanding of negation, and that the latter was probably more crucial. But it seemed to me that another variable, belonging to a different level of analysis, might also be relevant: the way in which negatives are generally used in

language. And following Charles Morris (1946), I called this a pragmatic factor. It would be inappropriate to consider it in the next section which is concerned with systematic research on the semantics of negation. It was some time before I realised the connection between the two.

(3) *Semantics.* Two information processing models of negation were independently postulated around 1970 by Herbert Clark and his associates at Stanford (Chase and Clark, 1972; Clark, 1972), and by Tom Trabasso and his associates at Princeton (Trabasso, 1970, 1971; Trabasso, Rollins and Shaughnessy, 1971). They represent at a more "molecular" level the mental operations which we have described, and they have both received a measure of empirical support which none of us had achieved. The models differ more in their conceptualisation than in their predictions, and they both follow in the traditions of McMahon (1963), Slobin (1963, 1966) and Gough (1965), in the sense that they are based on true-false verification tasks. Both investigators, however, each distinguish two types of model. Clark's "true" model and Trabasso's "response change" model assume what I have described as the mental deletion of "not", followed by an inversion of response. On the other hand, Clark's "conversion" model and Trabasso's "optional recoding" model correspond to changing the predicate of the sentence into affirmative form, e.g. reading "not even" as "odd". Trabasso has established that the responses tend to conform to this latter model (a) when the predicates are binary, or contradictories (rather than contraries), and (b) when the "object" is presented before the "description". Furthermore, this model predicts, not the interaction, but the ordering of the response times as follows: TA<FA<TN<FN. This has been independently confirmed by Clark (1972). We are inclined to agree with Clark when he calls this model a "cheating method" because of the reliance it puts on contradictory predicates, and since we are concerned with the interpretation of the interaction we shall ignore it. In other respects, Clark and his associates put more emphasis on linguistic analysis as a source of explanatory concepts. In

particular, they invoke the notions of deep structure, supposition and the "principle of congruence" (Clark, 1969). Clark has attempted to extend his model into a more general theory of comprehension, while Trabasso has extended his to explain the processes involved in the learning of a variety of logical connectives and rules. It would clearly be redundant to describe both models, and indeed it is difficult to do full justice to either. Instead, we shall restrict ourselves to a partial consideration of Clark's "true" model of negation.

It is postulated that subjects initially have available a potential response, set at "true" in a "truth index". The process of understanding is then assumed to occur in four stages. (1) The subject encodes the sentence in a mental representation. (2) The picture is encoded in the same representation. (3) The representations are compared to see whether they match. (4) An output occurs, and is converted into some form of response, usually "true" or "false". A consideration of just the true negative (TN) and the false negative (FN) is pertinent to understanding this conceptualisation of the interaction.

With TN, let us assume that the subject first reads the sentence, "B isn't above A", which he represents at stage (1) as "(false (B above A))". At stage (2) he then sees the picture of A above B which is represented as "A above B". At stage (3) the embedded strings "(false (*B above A*))" and "(true (*A above B*))" are compared, and found not to match. Hence the truth index is changed from "true", at which it was originally set, to "false". A third operation within this stage compares the embedding strings, "(*false* (B above A))" and "(*true* (A above B))", and finds that they too do not match. Hence the truth index is turned back again from "false" to "true". At stage (4) this output is converted into the response "true" — which is correct. This whole process has received substantial confirmation. As Clark (1972) puts it: "Subjects report changing their answer from true to false, and back again on True Negatives." With FN, on the other hand, the embedded strings, representing sentence and picture, are found to match at stage (3): "(false (*A above B*))" and "(true

(*A above B*))". Hence there is no need to change the truth index *initially* from "true" to "false", and an increment of time which this operation would require is not consumed.

We cannot consider here all the experiments conducted by Clark and his associates. Perhaps the most relevant is one by Young and Chase (cited in Clark, 1972), which showed that instructions to code the sentences and pictures in different ways confirmed exactly the predictions generated by the "true" and "conversion" models.

Further experiments tested the effects of "implicit negatives" which have a different underlying structure from "explicit" negatives. Clark makes considerable use of what he calls the "scope" of negation, which is derived in part from Klima's (1964) distinction between sentential and consistuent negation. In one study by Clark and Young (cited in Clark 1972) the sentence, "a plus isn't present" (explicit sentential negation) is compared with the apparently synonymous sentence, "a plus is absent" (implicit constituent negation). But, "a plus isn't present", is assumed to have the deep structure, "It is false to suppose that a plus is present", while "a plus is absent", is assumed to have the deep structure, "It is correct to suppose that it is false that a plus is present". In Clark's terms, the negative, "isn't present" contains the supposition, but the negative "is absent" is contained in the supposition. As predicted, the latter was verified more quickly than the former.

Finally, an experiment by Just and Carpenter (1971) studied the verification of three types of negative in relation to quantification: (1) explicit syntactic negatives, e.g. "None of the dots are red", (2) implicit syntactic negatives, e.g. "Few of the dots are red", and (3) semantic negatives referentially equivalent to (2), e.g. "A minority of the dots are red." All three types took longer to verify than their correlative affirmatives, but the semantic negatives behaved in rather different ways. Indeed, an unpredicted finding was that, under instructions to code the smaller subset, semantic negatives had a *shorter* reaction time than their correlative affirmatives. This was not the case with syntactic negatives. The conceptual

status of semantic negatives seems a little arbitrary, and certainly referential equivalence *per se* is not a satisfactory criterion. It would make the sentence, "The door is open", a semantic negative because it is referentially equivalent to the syntactic negative, "The door isn't closed". Such a conceptual point does not, of course, invalidate an investigation which poses a number of theoretical problems.

These three studies illustrate the range of linguistic hypotheses which derive from Clark's interpretation of negation. And yet I wonder whether, in striving to confirm precise, quantitative predictions, an important point about negation has been over-looked. We shall consider this in the next section, and it will take us back a little way in time.

(4) *Pragmatics.* My response to the studies of McMahon (1963) and Slobin (1963, 1966) was to investigate whether usage would affect the response to negation, over and above syntactic and semantic factors. Strawson (1952 p. 18) had pointed out that: "The standard and primary use of 'not' is specifically to contradict or correct; to cancel a suggestion of one's own or another's." In other words, negation functions within an affirmative context. A "thought experiment", which for technical reasons is difficult to control, illustrates this point. Compare in your mind the process of verifying the following sentences:

(1) 5 is not even
(2) 4 is even and 7 is not even

Most people find that (2) involves no detectable mental deletions, but (1) may lead to all the difficulties which we have considered.

In the exploratory experiment, conducted by Susan Carey, I distinguished two plausible contexts which might facilitate the process of denial (Wason, 1965). The "exceptionality hypothesis" stated, in effect, that it is more plausible to deny that an exceptional item possesses an attribute which makes it an exception, than to deny that any unexceptional item possesses

the discrepant attribute of the exceptional item. For example, given eight numbered circles, all of which are blue except for the third which is red, it will be easier to say, "Circle No. 3 is not blue" than "Circle No. 5 is not red", and that this effect will be less marked for the correlative affirmatives. The "ratio hypothesis" stated, in effect, that given two sets of stimuli, which differ in magnitude, it will be easier to deny that the smaller set possesses the property of the larger set than to deny the converse. For example, suppose the circles are unnumbered, it will be easier to say, "Exactly one circle is not blue" than "Exactly seven circles are not red."

The results confirmed the exceptionality hypothesis but not the ratio hypothesis. I attributed this (after several invaluable discussions with George Miller, and several abortive drafts of the paper) to the fact that the exceptional and unexceptional items had been explicitly related in the initial coding of the stimuli, e.g. "Circle No. 3 is red and the rest are blue". But under the ratio condition the two classes had initially been explicitly coded as independent and unrelated, e.g. "Seven circles are blue and one is red".

The paper has been criticised by Wales and Grieve (1969), (see Greene and Wason, 1970, for a rejoinder), Trabasso *et. al* (1971), Clark (1972) and Just and Carpenter (1971). Not all these criticisms can be correct because some are mutually exclusive. And yet the paper certainly is open to criticism. My own would be rather different: it would be that I was still too much tied to the "sentence-picture" paradigm, and that I was not sufficiently clear about the implications of my predictions when the experiment was performed. It mainly served to pose a new question. And it is worth remarking that, as a sign of the changing climate of opinion, when I presented it to a meeting of the Experimental Psychology Society in 1964, it was met by a barrage of stimulating, technical criticism about the validity of subtracting reaction times. But when the results of the first experiment (Wason, 1959) were presented to the same group in 1958, the paper met with a few perfunctory comments and a benevolent silence.

In 1964 my associates in London decided to widen the prob-

lem in different ways. For instance, Philip Johnson-Laird (1967) turned the problem round, and instructed the subjects to create a context which seemed appropriate for a negative utterance. Two rectangles had to be completely coloured in, with only red and blue crayons available, in order to depict a stimulus which might have been seen when each of the following sentences had been uttered:

(1) The left hand end of the strip is red
(2) The left hand end of the strip is not blue.

Within the task the two sentences are synonymous, but the prediction was confirmed that there would be more blue (and hence less red) in the drawing coresponding to (2) than to (1) because in (2) "blueness" is both emphasised and negated.

In rather the same way, Elizabeth Cornish (1971) instructed her subjects to complete the sentence:

The circle is not all ...

by pressing a key marked "blue", or a key marked "red", as soon as a circle was exposed in which the relative extent of these two colours was systematically varied over four stimulus conditions. It was predicted (1) that more responses corresponding to a particular colour would be made, when there was more of that colour in the circle, and (2) that such responses would systematically increase as a positive function of its relative extent. There was overwhelming support for (1), and considerable confirmation for (2). Both studies suggested that a contrast class modifies the interpretation of a denial even in a formal laboratory task.

Judith Greene (1970, a, b), however, was the first to have the insight that negation holds between two implicit, or explicit, sentences rather than between a sentence and a state of affairs. She argued that the "natural" function of a negative was to "signal a change of meaning", and that it performed an "unnatural" function when it preserved meaning. In her exploratory experiment, pairs of abstract sentences were printed on cards, and the task was to determine whether each pair had the same, or a different, meaning. The following pair

illustrates a "natural" negative because it differs in meaning from its affirmative partner, whatever two numbers are consistently substituted for x and y:

(1) x exceeds y
(2) x does not exceed y

The following pair illustrates an "unnatural" negative because it has the same meaning as its affirmative partner:

(1) y exceeds x
(2) x does not exceed y

With affirmative sentences (one active and one passive), on the other hand, the converse was predicted — sentences with the same meaning exert a "natural" function with respect to each other:

(1) x exceeds y
(2) y is exceeded by x

But the following pair have a different meaning, and hence exert an "unnatural" function with respect to each other:

(1) y exceeds x
(2) y is exceeded by x

The hypothesis that "natural" sentence pairs would be more rapidly discriminated, as "same" or "different", than "unnatural" pairs was overwhelmingly confirmed, every one of the subjects performing in accordance with the prediction. Subsequent experiments showed that the effect was primarily due to the difference between the two kinds of negative.

It was Greene's results which first illuminated for me the problematic affirmation-negation — truth-falsity interaction. Our "false negatives" corresponded to her "natural negatives": they both have a different meaning from the encoded description of a state of affairs, or from an explicit affirmative. In

Greene's case, of course, the sentences have no truth value (except in one experiment which will not be discussed), but if they were given an interpretation, it follows that they would be either both true, both false, or one true and the other false. Her innovation was to dispense with truth value, and yet still abstract the denial function of negation.

Negatives, when they are used in language, are usually false rather than true. They are, of course, generally true in relation to states of affairs, but I shall argue that states of affairs are not the appropriate criteria against which to assess the specific semantic role which they play. If I were to say, "The train wasn't late this morning", then in one sense the statement could count as a true negative, but that overlooks the reason for saying it. What the sentence does is to falsify the preconception of my listeners ("His train is always late"). And in this sense the statement is a false negative. It is just because of this that we do not, in talking, reading or thinking, seem to engage in computer-like processes, when we understand, or utter, negative sentences. In other words, as Johnson-Laird and I (Wason and Johnson-Laird, 1972) have put in: we do not have to recover the preconception because it has already occurred. (The real life corollary of the true negative is not, of course, a lie but a vacuous statement, e.g. "The train wasn't late this morning", when nobody expected it would be late.) It is not only experimental psychologists who have missed this point. In 1962 I discovered that Bertrand Russell, in the security of his arm-chair, had formulated the process of negation as follows:

> "When I say truly 'this is not blue' there is, on the subjective side, consideration of 'this is blue', followed by rejection, while on the objective side there is some colour differing from blue." (Russell, 1948).

He would have been the best subject of all to confirm Clark's "true" model of negation. It is the kind of account which one might give if one were to think about negation without considering the function it generally performs in a natural lan-

guage. It is a logician's description of negation as it operates within a calculus.

More recently, we have extended our investigations of negation (or more strictly falsification) into reasoning tasks. These present a very deep difficulty when a meta-inference has to be made, in spite of the structural simplicity of the problem (e.g. Wason, 1971). But even in direct inference tasks there is difficulty enough. For instance, Jonathan Evans (1972) showed that the *modus tollens* inference, which is known to be relatively difficult, is fairly easy for students, e.g.

(1) If the letter is A, then the number is 3
(2) The number is not 3
∴ The letter is not A

However, when the antecedent is negated, i.e. "If the letter is *not A*, then the number is 3", "The number is not 3", it was about three times as difficult to draw the valid conclusion, "The letter is A". In fact, the common response was to say that "nothing followed". On the other hand, a negated consequent in the conditional premise made no difference. There is evidently something difficult about denying a negative. But it could be argued that this is due either to "double negation", or to a failure to grasp that a negative can be falsified, and an affirmative elicited from it. The difficulty could be computational or conceptual.

Johnson-Laird and Tridgell (1972) have very recently simplified this task by substituting a disjunctive for a conditional sentence so that a valid inference can be made by falsifying one of its components. The effect is striking — all the more so because the correct response remains invariant under all experimental conditions. When the first premise is a disjunction, and the second one a categorical negative, the inference is relatively easy:

(1) Either John is intelligent or he is rich
(2) John is not rich
∴ John is intelligent

But when the negative occurs in the disjunction, and the second premise is a categorical affirmative, it is about twice as difficult to make the same inference:

(1) Either John is intelligent or he is not rich
(2) John is rich
*** John is intelligent

In the second version, Johnson-Laird remarks that there may be a tendency for subjects to argue that John is *not* not rich; it follows that he is rich. But this is the premise from which the argument started. "The whole of this double negation inference may then start again, and continue in an almost hypnotic fashion until the subject breaks the circle by concluding that nothing follows." It was also shown that when the explicit negative occured in the categorical premise, it was responded to faster than when a synonymous affirmative ("poor") was substituted for it. On the basis of these results the assumptions underlying the Clark and Trabasso models are criticised: "In order to grasp that one sentence denies another, the aim should be to establish *not* a one-to-one correspondence between them but a mutual inconsistency."

These results seem to me a beautiful corroboration of Greene's thesis that negatives cause little difficulty when they perform their natural function of denial. Clark and Trabasso might argue that all these experiments are no more than demonstrations which can be accommodated within their models. The assumption need only be made that a "natural" negative corresponds to the false, rather than the true, negative; and that the preconception corresponds to the encoding of the picture. In this sense, our studies could be said to show only how the coding processes, postulated by the information-processing models, might be realised in continuous discourse. But in a more general sense, it could be argued that the models are misleading, rather than wrong, because they neglect a context which is always present outside the psychological laboratory. Just as a philosopher might mistake Russell's account of negation for a description of the

mental processes which underly its use, so the psychologist might mistake the "true negative" for the norm. The difference is perhaps one of emphasis and interpretation, but it is an important difference because a person, who first encounters the models, might well be puzzled by the discrepancy between the theory and his own experience.

The proponents of the models might reply that such an interpretation is irrelevant to their aim of mapping the responses to the formal relations which hold between affirmation and negation, and truth and falsity, under controlled conditions. But these relations are not just formal; they are intimately connected with the way in which we think and talk. In the negation models the "true negative" does not conform to the ordinary usage of a natural language: it maintains truth rather than denies what is supposed to be true. It does duty for a redundant affirmative. This does not in the least detract from the surprising theoretical results obtained with the verification task. It suggests that the verification task, with its outputs "true" and "false", is an instrument which distorts our grasp of negation.

Conclusion

The future study of negation calls for a variety of different types of task which reflect the variety of functions which language performs, not only in describing objects, but in reasoning, writing, dialogue and monologue etc. Oldfield and Marshall (1968) have succintly summarised the long-term implication of pragmatic investigations of negation: "How is a particular way of saying something appropriate to the content one wishes to convey ? Problems of this nature take us beyond the study of the linguistic code *per se* and ultimately into studies of interaction between the belief systems of language-users".

No single technique is likely to solve these problems, or even enable us to formulate them clearly.

I am indebted to my research students who helped me think in different ways, and to all those who have worked on this problem, and so generously acknowledged my own contribution. I owe a special debt to my colleague, Dr. P. N. Johnson-Laird. Many of the ideas in this paper are derived from our joint collaboration on a forth-coming book. I should also like to thank the members of the Negation Project at the University of Gent, especially Professor Leo Apostel and Dr. Marc de Mey, for inviting me to address them, and for much stimulating discussion.

REFERENCES

BRUNER, J. S., GOODNOW, J. J., and AUSTIN, G. A., *A study of thinking*, New York: Wiley, 1956.

CHASE, W. G. and CLARK, H. H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory*, New York: Wiley, 1972.

CHOMSKY, N., *Syntactic structures*, The Hague: Mouton, 1957.

CLARK, H. H., Linguistic processes in deductive reasoning. Psychological Review, 1969, *76*, 387-404.

CLARK, H. H., Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics, Vol. 12: Linguistics and adjacent arts and sciences*, The Hague: Mouton, 1972 (in press).

CORNISH, E. R., Pragmatic aspects of negation in sentence evaluation and completion tasks. British Journal of Psychology, 1971, *62*, 505-11.

EVANS, J. St., B. T. Reasoning with negatives. British Journal of Psychology, 1972, *63*, 213-19.

GOLDMAN EISLER, F., *Psycholinguistics — experiments in spontaneous speech*, London: Academic Press, 1968.

GOLDMAN EISLER, F. and COHEN, M., Is N, P and PN difficultly a valid, criterion of transformational operations ? Journal of Verbal Learning and Verbal Behavior, 1970, *9*, 161-66.

GOUGH, P. B., Grammatical transformations and speed of understanding. Journal of Verbal Learning and Verbal Behavior, 1965, *4*, 107-11.

GREENE, J. M., The semantic function of negatives and passives. British Journal of Psychology, 1970 a, *61*, 17-22.

GREENE, J. M., Syntactic form and semantic function. Quarterly Journal of Experimental Psychology 1970 b, *22*, 14-27.

GREENE, J. M. and WASON, P. C., Negation: a rejoinder to Wales and Grieve. Perception and Psychophysics, 1970, *8*, 238-39.

HOVLAND, C. L. and WEISS, W., Transmission of information concerning concepts through positive and negative instances. Journal of Experimental Psychology, 1953, *45*, 175-82.

JOHNSON-LAIRD, P. N., An experimental investigation into one pragmatic

factor governing the use of the English language. Unpublished University of London Ph. D. thesis, 1967.

JOHNSON-LAIRD, P. N. and TRIDGELL, J. M., When negation is easier than affirmation. Quarterly Journal of Experimental Psychology, 1972, *24*, 87-91.

JONES, S., *Design of Instruction.* London: Her Majesty's Stationery Office, 1968.

JUST, M. A. and CARPENTER, P. A., Comprehension of negation and quantification. Journal of Verbal Learning and Verbal Behavior, 1971, *10*, 244-53.

KLIMA, E. S., Negation in English. In J. A. Fodor and J. J. Katz (Eds.), *The structure of language: Readings in the philosophy of language.* Englewood Cliffs, N. J.: Prentice Hall, 1964.

MCMAHON, L. E., Grammatical analysis as part of understanding a sentence. Unpublished University of Harvard Ph.D. thesis, 1963.

MILLER, G. A., *Language and communication*, New York: McGraw Hill, 1951.

MILLER, G. A., Some psychological studies of grammar. American Psychologist, 1962, *17*, 748-62.

MILLER, G. A. and MCKEAN, K. O., A chronometric study of some relations between sentences. Quarterly Journal of Experimental Psychology, 1964, *16*, 297-308.

MILLER, G. A. and MCNEILL, D., Psycholinguistics. In G. Lindzay and E. Aronson (Eds.) *The handbook of social psychology*, second edition, Vol. 3, Reading, Mass.: Addison-Wesley.

MORRIS, C. W., *Signs, language and behavior*, New York: Prentice-Hall, 1946.

OLDFIELD, R. C. and MARSHALL, J. C. (Eds.), *Language*, Harmondsworth: Penguin, 1968.

REICH, B., Affective constraints on the generating of speech. Unpublished University of London Ph. D. thesis, 1970.

RUSSELL, B., *Human knowledge: its scope and limits*, London: Allen and Unwin, 1948.

SLOBIN, D. I., Grammatical transformations in childhood and adulthood. Unpublished University of Harvard Ph.D. thesis, 1963.

SLOBIN, D. I., Grammatical transformations in childhood and adulthood. Journal of Verbal Learning and Verbal Behavior, 1966, *5*, 219-27.

SMOKE, K. L., Negative instances in concept learning. Journal of Experimental Psychology, 1933, *16*, 583-88.

STRAWSON, P. F., *Introduction to logical theory*, London: Methuen, 1952.

TRABASSO, T., Reasoning and the processing of negative information. Invited Address, American Psychological Association, 1970.

TRABASSO, T., Mental operations in language comprehension. Paper presented at "Language comprehension and the acquisition of knowledge", National Academy of Education and National Academy of Sciences, National Research Council, 1971.

TRABASSO, T., ROLLINS, H., and SHAUGHNESSY, E., Storage and verification stages in processing concepts. Cognitive Psychology, 1971, *2*, 239-89.

WALES, R. J. and GRIEVE, R., What is so difficult about negation ? Perception and Psychophysics, 1969, *6*, 327-32.

WASON, P. C., The processing of positive and negative information. Quarterly Journal of Experimental Psychology, 1959, *11*, 92-107.

WASON, P. C., Response to affirmative and negative binary statements. British Journal of Psychology, 1961, *52*, 133-42.

WASON, P. C., *Psychological aspects of negation*, London: Communication Research Centre, University College London, 1962.

WASON, P. C., The contexts of plausible denial. Journal of Verbal Learning and Verbal Behavior, 1965, *4*, 7-11. Reprinted in *Language*, R. C. Oldfield and J. C. Marshall (Eds.), Harmondsworth: Penguin, 1968.

WASON, P. C., Problem solving and reasoning. In *Cognitive Psychology*, A. Summerfield (Ed.), 1971, British Medical Bulletin, *27*, 206-210.

WASON, P. C. and JONES, S., Negatives: denotation and connotation. British Journal of Psychology, 1963, *54*, 299-307.

WASON, P. C. and JOHNSON-LAIRD, P. N., *Psychology of reasoning: structure and content*, London: Batsford, 1972.

P. C. WASON
Psycholinguistics Research Unit
University College London