

THE SCIENTIFIC STATUS OF
PROPAGANDA ANALYSIS

Alexander L. George

P-616

December 15, 1954

The **RAND** *Corporation*

1700 MAIN ST. • SANTA MONICA • CALIFORNIA

This paper has been prepared for the Work Conference on Theory and Technique of Content Analysis, to be held in February, 1955, at the University of Illinois, under the sponsorship of the Social Science Research Council's Committee on Linguistics and Psychology.

SUMMARY

When content analysis is used in order to infer antecedent conditions of communication, as is the case in propaganda analysis, logical operations are employed which resemble those necessary to scientific explanation. However, inferences from content analysis do not easily meet the following several requirements of scientific explanation: that general laws be used as part of the explanation; that the phenomenon to be explained be precisely designated; that all relevant antecedent conditions be identified; that the explanation (antecedent conditions and general laws) should bear a logical relationship to the phenomenon explained. In one form in which it has been applied, content analysis appears to depart further from scientific inquiry in that it deals with the actual words of the communication and not with a characterization of the communication arrived at by the analyst.

During World War II the technique of content analysis was developed and adapted for purposes of propaganda analysis of totalitarian communications. A group of analysts in a now defunct unit of the Federal Communications Commission (F.C.C.) attempted to make inferences about propaganda strategy and its underlying policy calculations from a close inspection of German radio and press communications. The work of the F.C.C. analysts has been studied intensively in order to reconstruct and codify the methods of inference which they applied with considerable success. The results of some of this research have been or will be reported elsewhere.

The present paper attempts to evaluate the scientific status of this particular application of content analysis. It is hoped that the results of the evaluation will be useful to those who are interested in advancing research methodology for the scientific study of communication and in developing other policy uses of the content analysis technique. Familiarity with the experience of the F.C.C. analysts may enable members of the larger scientific community to assist in improving the scientific status and policy usefulness of the content analysis technique.

A. REQUIREMENTS OF SCIENTIFIC EXPLANATION¹

The basic pattern of scientific explanation may be briefly described as follows:

¹ The present discussion is based upon statements of the logical empiricists. Technical language is avoided here in

Every explanation is composed of two major constituents:
(1) the phenomenon explained, and (2) the explanation for it. The explanation contains two parts: (a) specific antecedent conditions, and (b) general laws.

Any proposed explanation must satisfy certain logical and empirical conditions if it is to be adequate.

There are three logical conditions of adequacy:

(L1) The phenomenon explained must be the logical consequence of the explanatory statements (antecedent conditions and general laws)--i.e., it must be logically deducible from information contained in them.

(L2) The explanation must include general laws, and these must be actually required for the derivation of the phenomenon explained.

(L3) The explanation must have empirical content; i.e., it must be susceptible, at least in principle, to test by experiment or observation.

In addition there is one empirical condition of adequacy:

(E1) The explanation must be true. That is, the specific antecedent conditions and the general laws must be at least highly confirmed by all the relevant evidence available.

favor of a terminology closely related to content analysis. The statement made by C. G. Hempel and P. Oppenheim, "Studies in the Logic of Explanation," Philosophy of Science, Vol. 15, No. 2 (April, 1948), pp. 135-175, has been found particularly useful. (Cf. especially "Part I, Elementary Survey of Scientific Explanation.") For a fuller bibliography see that listed by Hempel and Oppenheim, op.cit., and, e.g., H. Feigl and W. Sellars, Readings in Philosophical Analysis (Appleton-Century-Crofts, Inc., New York, 1949).

B. COMPARISON OF CONTENT ANALYSIS INFERENCE WITH REQUIREMENTS OF SCIENTIFIC EXPLANATION

In scientific explanation, as has been noted, an attempt is made to relate the following: the phenomenon to be explained (P), antecedent conditions (A), and general laws (G). These three elements have the following counterparts in content analysis inference:

P = the content description; the content feature designated by the analyst to serve as an "indicator" for an inference may concern the syntactical characteristics of signs or their semantical components.² (Most of the work of the F.C.C. analysts involved semantical content analysis; that is, it dealt with the meaning of words as used by the propagandist.)

G = a generalized statement which relates content characteristics to some antecedent condition(s).

A = the "conditions" of the designated content characteristics, i.e., the antecedent conditions which have shaped the particular content characteristics singled out as indicators by the researcher.

If we may oversimplify a bit at this point, content analysis studies which attempt to make inferences begin with P and, utilizing some G, infer A. Content analysis studies, therefore, resemble scientific explanation rather than "scientific inquiry" or "scientific prediction."³

To what extent do content analysis inference studies satisfy scientific requirements?

² For elaboration, see Irving L. Janis, "Meaning and the Study of Symbolic Behavior," Psychiatry, Vol. 6, No. 4 (November 1943), pp. 425-440.

³ Scientific Inquiry usually (though not always) begins with P and A and induces G. (Some content analysis studies are designed to produce generalizations, linking P to A, which will help support inferences on other occasions.) Scientific Prediction begins with G and A and proceeds to P. (In some cases content analysis attempts to predict what P will be.)

(1) Requirement of Empirical Content: Antecedent conditions inferred in content analysis do meet the requirement that a scientific explanation have empirical content so that it be susceptible, at least in principle, to test by experiment or observation (Requirement L3, p. 2). The antecedent conditions of interest in propaganda analysis usually fall into three general types: actual events, elite policy calculations, and propaganda strategy, all of which are obviously subject to observation, at least in principle. The difficulty is not this but, rather, the practical one of obtaining direct evidence of the conditions inferred that will serve for purposes of validation.

Each inference, when it is made, is obviously conjectural in that it lacks validation. This situation is inherent in the use of content analysis as an inferential technique. For, if empirical data on the antecedent condition of interest were available at the time the inference is made, there would be no point to using content analysis (or, for that matter, any other indirect technique) to infer that antecedent condition. A problem exists only because the condition in question is not directly and empirically known.

In some types of inquiry, the empirical data required to confirm an explanatory hypothesis or an inference may be available shortly thereafter. This possibility usually does not exist in the propaganda analysis application of content analysis. As a result, it is difficult to improve the quality of a propaganda analysis operation by means of a timely

systematic assessment of the successes and failures which it is achieving. Given the nature of interstate relations, direct empirical data on the conditions inferred becomes available, if at all, only after it has ceased to have any practical importance for the policy research operation, i.e., after military defeat in war, after a radical or revolutionary change in government, or simply through the passage of time. It was only after World War II, for example, that such data (official Nazi records, diaries, interrogations of captured personnel, etc.) became available for validating many of the inferences made by the F.C.C. and other research agencies.

In the absence of empirical data for direct verification, indirect attempts at verification may be made. One familiar method is to use the course of subsequent historical events as a basis for deciding whether a previous inference was correct. This approach contains many possibilities of error, especially for the types of inferences made in policy research, and cannot be considered a satisfactory basis for indirect verification. Thus, for example, the inference that certain content characteristics indicate a desire for diplomatic rapprochement with an opponent is not necessarily verified by the fact that several months later such a rapprochement takes place. The desire may have developed after the inference was made; or, the inference, though "correct," may have been based upon the wrong content indicators, or upon an incorrect reconstruction of propaganda strategy. Similarly, subsequent events which appear to disprove an inference may not do so at all. Thus, the fact that no steps were actually taken to

bring about a diplomatic rapprochement may mean simply that the original intention in this direction was abandoned; it does not necessarily mean that this intention was not entertained at the time the inference was made by the content analyst.

Just because empirical validation of results is difficult and remote in time, content analysis inference tends to be regarded as a highly hypothetical type of inquiry as compared to other forms of scientific explanation. Indeed, an element of uncertainty always attends the use of content analysis in policy research of this type. Both content analysts and their clients would be happier if some of the inferences could be conclusively verified in the course of the policy research operation. (This limitation, however, to some extent also characterizes other methods employed for obtaining information about aspects of events, elite behavior, and strategy which are not susceptible to direct observation.)

(2) Requirement of General Laws: When engaged in inference, content analysts usually have considerable difficulty in attempting to satisfy the requirement that general laws be used (Requirement L2, p. 2). The mere statement of an inferential hypothesis asserting a relationship between content characteristics (P) and antecedent conditions (A) clearly does not meet this requirement and is not an adequate scientific explanation of the phenomenon.

The difficulty in question is not surprising, given the rudimentary nature of present knowledge about political

communication. Available generalizations about political communication frequently lack concrete relevance for purposes of inference; their usefulness is often restricted also by their imprecise formulation and uncertain validity. As a result, generalizations tend to be adduced ad hoc by the content analyst and, more often than not, are only implicit in the reasoning by means of which he supports an inference. Moreover, the content analyst himself, consciously or unconsciously, seems to attribute low explanatory power to such generalizations and does not consider that they give strong support to his inferences.⁴

As a result, it is sometimes argued that the requirement in question be waived in content analysis on the ground that it is not always possible or desirable to replace(!) good judgment by rules. This argument may be ~~pr~~aphrased as follows: In content analysis, as in other forms of expert guessing, it is possible to make sound inferences from particular to particular without being able to formulate explicitly any acceptable general law or supporting generalization. The fact that no such rule is invoked does not necessarily mean that the inference is incorrect or unsound. For the correctness or incorrectness of the inference is obviously a matter for empirical test, whether or not the generalization and reasoning on which it is based can be formulated. Thus, for example, the expert poker player can usually tell whether his opponent is bluffing; he makes use of generalizations for this purpose but would be hard put to

⁴ It may be noted that, at the same time, content analysts recognize the need for developing generalizations relevant for purposes of inference by means of special content analysis studies.

articulate the basis for his judgment. If pressed, he might succeed in listing some criteria which he employs in making such an inference, but the inadequacy of these criteria might well be exposed if some other person attempted to apply them. Therefore, the conclusion is drawn that perhaps in content analysis, as in poker and other forms of inference and diagnosis, "expertness" must remain intuitive. Although the expert makes use of generalizations privately, he cannot articulate them into a set of formal rules.

This general argument, familiar in many other types of inquiry as well, is usually rejected by students of science. In addition, there are practical policy reasons for asking the content analyst to present relevant reasoning in support of his inferences. The analogy with the expert poker player is not an apt one for several reasons. In the first place, poker has a well-defined statistical basis, which propaganda analysis does not. Expertness in poker is possible, but it is not certain that comparable skill in propaganda analysis can be achieved. Moreover, in poker it is possible to validate a hunch (inference) immediately after each hand as to whether the opponent is bluffing or as to the relationship between his betting strategy and his cards; this is not the case in content analysis inference, as has already been observed. The value of the empirical test in poker is plainly revealed in the rule that you must pay to see what the opponent holds. Under the favorable conditions of poker, a player can improve his diagnosis of an opponent's

method of play within a relatively short time by directly observing how he plays different hands; and he can apply knowledge about the opponent thus obtained in following hands. This type of learning, as we have seen, is more difficult in propaganda analysis. Finally, because poker is a game for individuals, the poker player is both an intelligence analyst and a decision-maker. In contrast, the content analyst is only an intelligence specialist and, therefore, has the difficult problem of demonstrating the plausibility of his inferences or his "expertness" to policy-makers.

For all these reasons, the requirement that inferential hypotheses be supported by relevant reasoning cannot be easily waived in policy applications of content analysis. An inference which is unsupported by explicit reasoning may be correct, but, if generalizations adequate for supporting the inference can be formulated, the possibility of improving the level of inference and teaching the skill to others is thereby increased. Since empirical validation of content analysis inferences is difficult, the determination of "expertness" on the basis of past performance is not an entirely feasible alternative. It becomes even more important, therefore, to have an internal check on the plausibility of individual inferences. The reasoning and generalizations cited by the content analyst in support of inferences provide this possibility.

(3) Precise Designation of Phenomenon To Be Explained:

The characteristics of the phenomenon to be explained (P) must be precisely designated in any type of scientific inquiry; it is particularly difficult to do so in content analysis. The propaganda materials examined by the content analyst are capable of yielding a large number of content designations. To a greater extent than in most forms of scientific explanation, the phenomenon to be explained is not given but must itself be identified and designated by the content analyst. The sensitivity and diagnostic skill of the analyst consist in spotting which aspects of the content (possible P) may be related to which antecedent conditions.⁵ Ambiguous designations of P have serious consequences, noted below.

(4) Logical Relationship of Explanation to Phenomenon Explained: The requirement that a scientific explanation (i.e., statement of antecedent conditions and general laws) should bear a logical relationship to the phenomenon explained may be considered to be a requirement also in content analysis. That is, P should be logically deducible from A and G (Requirement L1, p. 2). In content analysis one encounters various difficulties in attempting to meet this requirement:

- (a) supporting generalizations are not always fully adequate;
- (b) not all relevant antecedent conditions are specified;
- (c) the designation of P itself may be ambiguous, and, when this occurs, P is not clearly the logical consequence of

⁵ This problem is considered at some length in RAND P-617, "Quantitative and Qualitative Procedures in Content Analysis."

propagandist. Occasionally, it happens, and is to be expected, that the content analyst is fooled about some of the variables he has assumed to be stable.

C. THE PROBLEM OF GENERAL TERMS IN CONTENT DESCRIPTION

Modern science aims at the establishment of general laws of a causal or probabilistic character. Knowledge which fails to achieve a causal or probabilistic character, however useful it may be in everyday life or in specialized pursuits, is regarded as "pre-scientific."⁷

Scientific inquiry deals not with individual events themselves but with specified characteristics of events, or types of events. It is sometimes argued that events emerging from the activity of human beings, singly or in groups, have a uniqueness which makes them inaccessible to efforts at explanation by means of causal or probabilistic generalizations. This argument is based upon a misunderstanding of the logical character of scientific inquiry and explanation. It is true that scientific inquiry seeks for uniformities and, in seeking for causal or probabilistic laws, presupposes repeatability of phenomena.

But the phenomena of interest to science are not individual events, but events of a specified type. A causal or probabilistic law merely asserts that events with certain specified characteristics are accompanied, or tend to be accompanied, by other events with certain other specified characteristics. What is needed for the development and

⁷ See Patrick Gardiner, The Nature of Historical Explanation, London, Oxford University Press, 1952.

testing of such laws is the recurrence not of the same individual event, but of separate events which have the same specified characteristics. It is the repetition of types of events, not of an individual event, which is the focus of scientific inquiry. The "uniqueness" of events is not an exclusive feature of human history. In the physical sciences no less than in psychology and the social sciences, every individual event is unique in the sense that it never repeats itself with all of its characteristics.

It is for this reason that those who have attempted to develop content analysis as a technique for the scientific study of communication include the requirement that general terms, or categories, be used in describing content. This requirement receives perhaps its most explicit formulation in the definition of content analysis by N. C. Leites and I. de Sola Pool.⁸ In practice, this requirement means that the language into which a content-descriptive observation is cast should not be identical with, but more general than, the language of the content in question. (The relationship between selection of content categories and efforts to develop a systematic theory will not be considered here.)

The requirement of general terms of description has been closely associated with emphasis on quantitative content analysis. Analysts at the Federal Communications Commission, however, found that quantitative attributes of language

⁸ On Content Analysis, Document No. 26, Library of Congress, Experimental Division for the Study of Wartime Communication (September 1, 1942), pp. 1-2.

were of limited utility for purposes of inference. Instead, they came to depend to a considerable extent on qualitative attributes of content, whether or not certain content characteristics were present in the communications under inspection.

The requirement that general terms be used in content description was usually met in the "frequency" approach, but not in the "non-frequency" approach employed by the F.C.C. analysts.⁹ For example, the use of the word "counter-terror" by Goebbels in one of his speeches was singled out as significant for purposes of inference. In this and other examples of the non-frequency approach, no effort was made by the F.C.C. analyst to abstract or generalize certain characteristics of the words Goebbels actually used.

The question arises whether violation of the requirement that general terms be used is inherent in the use of non-frequency content description for purposes of inference. The F.C.C. analysts' use of the original language event itself, rather than some abstracted statement of some of its characteristics, may be explained on several grounds. First, since the immediate inference of interest apparently could be made without generalizing the content indicator, it may have seemed unnecessary for the analyst to do so. Secondly, since some effort would be required to generalize non-frequency

⁹ In reviewing the work of the F.C.C. the writer found it useful to replace and elaborate the usual distinction between "qualitative" and "quantitative" content analysis. A content-descriptive statement referring to the number of times a certain characteristic is present is defined as a frequency content indicator. A content-descriptive statement referring to the mere presence or absence of a certain characteristic is defined as a non-frequency content indicator.

content indicators, this task may have been pushed aside in the rather hectic work situation in which F.C.C. analysts operated. Thirdly, the F.C.C. analysts may have been reluctant to generalize non-frequency indicators for fear of losing thereby some of the meaning of the original language event that appeared necessary for making the inference. (As was noted above, the F.C.C. analysts depended largely on semantical components of propaganda for purposes of making inference.)

It is, of course, only the third of these explanations which raises the possibility that non-frequency content analysis may be unable to meet the requirement in question.

The nature of non-frequency content description and its role in inference need further study. It is possible that non-frequency indicators are subject to some form of qualitative measurement.¹⁰ But it is not clear whether available techniques for qualitative measurement are adequate for this purpose or whether increased precision resulting from such measurement would significantly improve the range and quality of inferences in this type of policy research.

When evaluated according to the formal requirements of modern scientific inquiry, the non-frequency approach probably must be regarded as a "pre-scientific" method of analysis. But it may not be a method which can be arbitrarily

¹⁰ See Paul F. Lazarsfeld and Allen H. Barton, "Qualitative Measurement in the Social Sciences," included in D. Lerner and H. D. Lasswell (eds.), The Policy Sciences, pp. 155-192.

and immediately cast into a fully scientific mold without reducing its policy usefulness.

The non-frequency approach invites attention also from a broader philosophical standpoint. For this approach may carry with it the assumption that at least some of the language of manipulative communications must be regarded as sui generis by the propaganda analyst in order that he may fully grasp the semantical component of the speaker's meaning and relate it to its pragmatic dimension. Such a view, if articulated clearly, would hold that--at least for such purposes--the semantical component of language is not subject to scientific study. That is, it could be argued that the language models useful for such purposes will be neither causal nor probabilistic in nature.¹¹

¹¹ The view that language models must be "interpretative," rather than causal or probabilistic, has been explicitly formulated and given broad application by Paul Kecskemeti in his Meaning, Communication and Value (Chicago: University of Chicago Press, 1952). For the possibility that behavioral models other than the causal or probabilistic may emerge from linguistics and from mathematics, see Clyde Kluckhohn, "The Study of Culture" (included in D. Lerner and H. D. Lasswell /eds./, The Policy Sciences, 1951), especially his remarks on the work of the anthropological linguist B. L. Whorf. See also Paul Kecskemeti, "The 'Policy Sciences': Aspiration and Outlook," World Politics, IV, No. 4 (July, 1952), especially pp. 526-533.

On the other hand, Irving L. Janis ("Meaning and The Study of Symbolic Behavior," op. cit., p.p.433-434) argues that it is possible to use a frequency or statistical criterion as a basis for constructing the semantical rules for a "natural" (non-scientific) language. He derives support for this view from the writings of Rudolf Carnap and other modern logicians. However, Janis concedes that use of a frequency criterion for this purpose in semantical content analysis would tend to rule out the recording of meanings which occur infrequently in the language habits of the communicator or source under investigation. Moreover, and of considerable

It should be noted that non-frequency content analysis was evidently developed by F.C.C. analysts in conjunction with a particular method of inference.¹² This "indirect" method appears to have required that an effort be made to grasp rather fully the semantical meaning of language employed by the propagandist. For this reason, the analysts seem to have preferred to deal with language phenomena as individual events, in order to explore in detail its communications and situational contexts.¹³

relevance to the present discussion of propaganda analysis, Janis (p. 438) believes it unlikely that many correlations between semantical and pragmatic characteristics of language will be discovered which will be high enough to permit semantical content characteristics to be used as precise (quantitative) indicators of pragmatic response. In other words, though Janis argues that valid semantical content description can be achieved by employing a probabilistic model for the semantical rules of the language, he does not appear to be optimistic regarding the usefulness of probabilistic behavioral models for the relationship between semantical and pragmatic components of language. In a sense, therefore, Janis' remarks--if we have interpreted them correctly--forecast the difficulty of employing a scientific form of semantical content description in propaganda analysis, later evident in the work of the F.C.C. (The study by Janis is one of the few available discussions of the problem of validity in semantical content description.)

¹² This has been codified by the writer and labelled the "indirect" method of inference. It consists, essentially, in taking into account the purposive character of propaganda communications. Thus, the first major step in the indirect method is always an inference about the propaganda goal of the communication under analysis. From this, further inferences are made as to the policy which that propaganda strategy is intended to advance, the estimates and forecasts on which that policy is based, and perhaps the events which lie behind the estimates and forecasts.

¹³ Related to this is the problem of achieving validity in descriptions of the semantic meaning intended by a speaker. The problem is dealt with in greater detail in RAND P-617, "Qualitative and Quantitative Procedures in Content Analysis."

It should be noted, however, that F.C.C. analysts also frequently dealt with the content of propaganda communications in such a way as to satisfy the requirement that general terms of description be used. This was the case, typically, when they utilized frequency content indicators for purposes of inference. The frequency approach was employed largely, but not exclusively, in conjunction with the "direct" method of inference.¹⁴ A major limitation of this method for purposes of producing inferences was the lack of appropriate generalizations linking types of content characteristics to types of antecedent conditions having to do with elite policy calculations.

¹⁴ In the "direct" method, inference is from content characteristics to elite policy calculations and related events; unlike the "indirect" method, therefore, the "direct" method bypasses the intervening variable, propaganda strategy. Its justification for doing so is that it attempts to locate content indicators which are "independent" of propaganda strategy, though related to policy calculations or events.