

# 3.1

## CONTINGENCY ANALYSIS

### *Validating Evidence and Process*

CHARLES OSGOOD\*

#### INFERENCES AND INDICATORS

Making an inference (or prediction) in content analysis involves at least the following: (1) some indicator or class of indicators that can be identified in the message sequence, (2) some state or process in the individuals producing or receiving the message, and (3) some dependency between these two such that the presence, absence, or degree of the former is correlated more than by chance with the presence, absence, or degree of the latter.\*\* The events in messages that *might* serve as indicators (correlate unspecified) are practically infinite—the frequency or locus of occurrences of the first-person singular pronoun “I,” the sheer magnitude or rate of output in word or other units, pitch and/or intensity oscillation of the voice in various message segments, the probability level of the syntactical alternatives chosen, and so on ad infinitum. Similarly, the

states of individuals that one *might* make inferences about (again, correlate unspecified) are as infinite as the classificatory ingenuity of all of the members of the American Psychological Association put together—the intelligence, communicative facility, or racial origin of the speaker, his anxiety, aggressive, or sexuality level, his association, attitude, or value structure, his semantic or formal language habits, and so on. Most (if not all) of the characteristics of an individual, in one way or another, probably influence what happens in his communications. But the rub lies in (3) above—some indicator having a non-chance relation to the characteristic in which we are interested must be isolated—and so far, psycholinguistics has had little more than suggestions or hunches to offer.

[Content analysts] . . . are likely to be most interested in *specific inferences*; for instance, does country A intend to attack country B and

\*From Osgood, C. K. (1959). The representational model and relevant research methods. In I. de Sola Pool (Ed.), *Trends in content analysis* (pp. 33–88). Urbana: University of Illinois Press. (Excerpt represents pp. 33–37, 54–71, and 73–77).

\*\*When this text-context relationship is operationalized for use in a content analysis, we call it an “analytical construct.” See Krippendorff, K. (2004:34–35, 171–187). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

when? Driven by internalized demands for scientific rigor, the academically oriented user of content analysis is likely to be most interested in *general inferences*; for instance, is there a general lawful relationship such that increase in the drive level of the speaker is accompanied by simplification and normalization of his semantic and structural choices? . . . [T]here is no necessary incompatibility here: just as the validation of many specific inferences by practically oriented users may provide insights into general relations, so the gradually accumulating generalities of the academician may enrich the inference base for the practical content analyst. The ideal situation is probably that in which “tool makers” and “tool users” work in close association.

Of all . . . source or receiver characteristics, which might be inferred from the content of their communications, . . . four (are outstanding. *Attention* or interest, inferred from the relative frequencies with which lexical items are produced; *Attitudes*, inferred from the use of evaluative terms; *Language correspondences* or linguistic habits, inferred from context dependent expectations; and *Association structures*, inferred from the contingencies between content items in a source’s messages, regardless of either frequency of usage or evaluation. This chapter concerns the latter.)

#### ASSOCIATIONS AND DISSOCIATIONS

An inference about the “association structure” of a source—what leads to what in its thinking—may be made from the contingencies (or co-occurrences of symbols) in the content of a message. This inference is largely independent of either “attention level” (frequency) or “evaluation” (valence). One of the earliest published examples of this type of content analysis is to be found in a paper by Baldwin (1942) in which the contingencies among content categories in the letters of a woman were analyzed and interpreted.

If there is any content analysis technique, which has a defensible psychological rationale,

it is the contingency method. It is anchored to the principles of association, which were noted by Aristotle, elaborated by the British Empiricists, and made an integral part of most modern learning theories. On such grounds, it seems reasonable to assume that *greater-than-chance* contingencies of items in messages would be indicative of *associations* in the thinking of the source. If, in the past experience of the source, events A and B (e.g., references to FOOD SUPPLY and to OCCUPIED COUNTRIES in the experience of Joseph Goebbels) have often occurred together, the subsequent occurrence of one of them should be a condition facilitating the occurrence of the other: the writing or speaking of one should tend to call forth thinking about and hence producing the other. It also seems reasonable to assume that *less-than-chance* contingencies of items in messages would be indicative of *dissociations* in the thinking of the source. If, in the experience of the source, events A and B (e.g., MOTHER and SEX in a psychotherapy case) have often been associated, but with fear or anxiety, the occurrence of one of them should lead to the inhibition of the other. Such inhibition might be either central (unconscious and involuntary) or peripheral (conscious and deliberate).

#### AN EXPERIMENTAL TEST OF THE BASIC ASSUMPTIONS<sup>1</sup>

In applying contingency analysis to real problems, such as propaganda study and psychotherapy, we would like to use the data about what things co-occur in messages to make inferences about a person’s association structure and also about what things have gone together in his (or her) experience; that is, about the experiential basis for his or her association structure. Unfortunately, however, in such application situations we seldom if ever have any data with which to validate our inferences. Usually we have only the messages produced, not the source who produced the messages (and who could give us other indices of his association structure) and certainly not

the history of his experience. In order to test the basic assumptions of this method, therefore, it is necessary to develop a controlled experimental situation in which (1) the experiential history can be approximately known, and (2) the association structure can be estimated independently of the message structure. The following experiment provides such conditions.

### Hypotheses and General Design

Our general assumption is that (1) *contingencies in experience* come to be represented in (2) *an individual's association structure* by patterns of association and dissociation of varying strengths, which help determine (3) *the contingencies in messages* produced by this individual. We require a simple situation in which we can measure

(1)  $F_a(\mathbf{b}) > F_a(\mathbf{c}) > F_a(\mathbf{d}) \dots > F_a(\mathbf{n})$  — the varying frequencies (F) in experience with which an event (a) is followed by other events (b, c, d . . . n);

(2)  $P_a(\mathbf{b}), P_a(\mathbf{c}), P_a(\mathbf{d}) \dots P_a(\mathbf{n})$  — the varying probabilities (P) with which subjects exposed to the above experience will associate items b, c, d . . . n when *some other person* (the experimenter) gives a, thus providing a measure of association structure (associational probability); and

(3)  $P_a^*(\mathbf{b}), P_a^*(\mathbf{c}), P_a^*(\mathbf{d}) \dots P_a^*(\mathbf{n})$  — the varying probabilities with which subjects exposed to the above experience will produce items b, c, d . . . n after *they themselves* have produced a. This provides a measure of message contingency (transitional probability). If we think of the subject in this experiment as a communicating unit in the information theory sense,  $F_a(\mathbf{b})$  is the input to the unit and  $P_a^*(\mathbf{b})$  is the output. The experimenter determines the input in such a way that  $F_a(\mathbf{b}) > F_a(\mathbf{c}) > F_a(\mathbf{d})$ . (It is assumed that before this experience the associations between these items are random across subjects, and materials for the experiment are selected to approximate this condition.) The theory, which we are testing, may be stated more formally as a series of hypotheses.

**Hypothesis I.** Exposure to a sequence of paired events such that  $F_a(\mathbf{b}) > F_a(\mathbf{c}) > F_a(\mathbf{d})$  will result in a non-chance association structure among these events such that  $P_a(\mathbf{b}) > P_a(\mathbf{c}) > P_a(\mathbf{d})$ .

**Hypothesis II.** Given an association structure such that  $P_a(\mathbf{b}) > P_a(\mathbf{c}) > P_a(\mathbf{d})$  in a set of subjects, sequential messages by these subjects limited to these events will display contingencies (transitional probabilities) such that  $P_a^*(\mathbf{b}) > P_a^*(\mathbf{c}) > P_a^*(\mathbf{d})$ .

**Hypothesis III.** Given exposure to a sequence of paired events such that  $F_a(\mathbf{b}) > F_a(\mathbf{c}) > F_a(\mathbf{d})$  and subsequent production of sequential messages limited to these events, message contingencies will be such that  $P_a^*(\mathbf{b}) > P_a^*(\mathbf{c}) > P_a^*(\mathbf{d})$ . This dependency relation between input and output assumes mediation via the subject's association structure.

**Hypothesis IV.** The dependency relation between association structure and message contingency (described in Hypothesis II) will be greater than the dependency relation between input contingency and message contingency (described in Hypothesis III). This derives from the assumption that message contingencies depend directly upon the association structure of the subject and only mediately upon his experience; to the extent that individual subjects have prior associative experience with the items, these associations will also influence the final structure.

**Hypothesis V.** The degree of dependence (1) of association structure upon experiential contingency and (2) of message contingency upon experiential contingency will be a direct function of the frequency of experiential contingency,  $F_a(\mathbf{b})$ . In other words, we assume that modification of association structure (and hence transitional message structure) varies with the frequency with which events are paired in experience—a straightforward psychological association principle. With respect to measurement, this implies that the more frequent pairings in experience have been,

the more significant will be the deviations of associational and transitional probabilities from chance.

**Hypothesis VI.** The degree of dependence between association structure and message contingency will be relatively independent of the frequency of experiential contingency. This assumes that whatever pre-experimental associations between items exist in individual subjects will determine both associative and transitional (message) contingencies; hence dependency relations here should be relatively independent of experimental inputs.

The burden of this analysis, if substantiated in the results, would be that contingency content analysis provides a valid index of the association patterns of the source, but only a mediate and tenuous index of his life history. It is realized, of course, that this “laboratory”

approach side-steps many of the problems that arise in practical applications of contingency analysis; some of these will be considered later under a critique of the method.

### Method

Two groups of 100 subjects each were shown 100 successive frames of a single-frame film strip. On each frame was a pair of girls' names, for example BEATRICE-LOUISE. There were only ten girls' names altogether, but these were so paired that (1) each name would appear equally often on the left and on the right, (2) the ordering of frames with respect to names was random, and (3)—the main experimental variable—each name appeared with others with different frequencies. The pattern of input pairing shown below for JOSEPHINE was duplicated for each of the ten names (with different specific names, of course):

JOSEPHINE-BEATRICE	6	LOUISE-JOSEPHINE	6
JOSEPHINE-CYNTHIA	3	GLADYS-JOSEPHINE	3
JOSEPHINE-HAZEL	1	ESTHER-JOSEPHINE	1
	with SARAH	0	
	with ISABELLE	0	
	with VALERIE	0	

Subjects were asked simply to familiarize themselves with the names. Following viewing of the 100 frames, two different measures were taken. (1) *Association test*. Each of the girls' names was shown separately on the screen for eight seconds, and subjects were instructed to write down the first other girl's name that occurred to them. Here the experimenter provides the stimulus—associative probability. (2) *Transitional contingency test*. Subjects were given little booklets and instructed to write one girl's name successively on each page, filling in as many pages as they could and not looking back. Here the stimulus for each response is the subject's own previous behavior.

The group that had the associational test first and the transitional last will be referred to as Group I; the one that had the transitional

test first and the associational last will be called Group II. Three tables were formed for each group. The *input table*, the same for both groups, gave the relative frequency (percent) with which each name had been paired with every other name on the presentation frames, without regard to the forward or backward direction of association. Since each name appeared on 20 frames, an item paired six times with another would have this noted as 30 percent of its appearances, three times, 15 percent, and one time, 5 percent. The *association table* gave, for each stimulus name, the relative frequency (percent) of subjects giving each of the ten possible response names. Thus, if 16 of the 100 subjects wrote BEATRICE when they saw JOSEPHINE, 16 per cent was entered in the appropriate cell. The *transitional table* gave, for each self-produced

stimulus name, the relative frequency (percent) of subjects giving each of the ten possible response names. Thus, if JOSEPHINE appeared in the booklets of 79 subjects and was followed immediately by the name HAZEL in the booklets of 11 of these subjects, 14 per cent was entered in the appropriate cell.<sup>2</sup>

## Results

Table 1 gives the correlations obtained among input, association structure, and message contingency. With regard to the first hypothesis, it can be seen that the  $r$  between input frequency and associative probability is .58 for Group I and .37 for Group II—these are both significantly greater than zero and in the expected direction. The fact that the correlation is considerably higher for Group I than Group II may reflect the effect of interpolating the transitional test and the consequent greater remoteness of the stimulus input from the act of producing the association in Group II.

Relations between input frequency and transitional (message) contingency are both in the predicted direction but are not significantly different from zero. Hypothesis III is thus not confirmed at a satisfactory level of significance. It is interesting to note, however, that (1) the input/transitional  $r$  is actually lower in Group II, where the transitional test immediately followed the input, than in Group I, and (2) the relation between input and transitional probabilities seems to vary with that between input and associational probabilities—as if (as hypothesized) the transitional contingencies depended upon the associative structure.

Regarding Hypothesis II, it may be seen that the  $r$  between associational and transitional probabilities is positive and significant for both groups. That the degree of relation between associational and transitional probabilities is approximately the same (particularly when the continuous raw data are correlated, .46 and .42) for both groups substantiates Hypothesis VI; as is shown, despite the gross differences between Groups I and II in degrees of correlation between input frequencies and both measures, the relation with association structure and transitional message structure is the same. With regard to Hypothesis IV, it can be seen that for both groups the correlations between associational probabilities and transitional probabilities are higher than those between input frequencies and transitional probabilities, as anticipated.

Finally, there is Hypothesis V—that the degree of dependency of both associational and transitional probability upon the input frequencies varies with the absolute frequency of input pairing. To test that we examine whether the predictability of a response name from knowing the stimulus name varies with the frequency of pairing in the input. For each subject the number of “correct” responses given in each frequency category was recorded. (A Dixon-Mood sign test was used to determine significance.) For all conditions except the backward direction of association on the transitional test, six pairings yielded significantly more “correct” associations than either three or one pairings. The lower frequencies of pairing, three and one, were not significantly different from each

**Table 1**

<i>Dependency Relations</i>	<i>Group I</i>	<i>Group II</i>
Input/Associational	.58 <sup>a</sup>	.37 <sup>a</sup>
Input/Transitional	.19	.09
Associational/Transitional <sup>b</sup>	.39 <sup>a</sup> (.46) <sup>a</sup>	.48 <sup>a</sup> (.42) <sup>a</sup>

a. Significantly greater than zero at the 5 percent level or better.

b. The bracketed values for associational/transitional in this table were computed from the continuous raw data prior to transformation into discrete stepwise values.

other or from zero pairing. As might be expected in a culture that reads from left to right, “forward” associations were significantly stronger than “backward” associations.

### Conclusion

This experiment was designed to test certain assumptions that seem to give value to a contingency method of content analysis: (1) that the association structure of a source depends upon the contingencies among events in his life experience, and (2) that inferences as to the association structure of a source can be made from the contingencies among items in the messages he produces. This experiment provided conditions in which the contingencies among events occurring to human “sources” could be at least partly manipulated and hence known. It also provided conditions in which the resultant association structures of these “sources” could be determined independently of the contingencies in the “messages” (transitional outputs) they produced. Both of the major assumptions above were supported by the results, association being shown to be dependent upon input contingencies and transitional output contingencies upon association structure to significant degrees. The results also indicate that whereas “message” contingencies are dependent upon association structure, they are only remotely dependent upon experienced input within the experiment itself; that is, non-chance associations between items existed prior to the experimental input manipulation and also influenced transitional contingencies. In general, the degree to which input influences both association structure and transitional contingency is a function of the frequency of input pairing.

### NATURE OF CONTINGENCY ANALYSIS

In the application of the contingency method as a kind of content analysis, in contrast to the experimental situation just described, we are limited to events in messages, and from them try to make inferences about the association

structure of their source. The message is first divided into units, according to some relevant criterion. The coder then notes for each unit the presence or absence of each content category for which he is coding. The contingencies or co-occurrences of categories in the same units are then computed and tested for significance against the null (chance) hypothesis. Finally, patterns of such greater-than- or less-than-chance contingencies may be analyzed. This may be done by a visual model, which gives simultaneous representation to all of the relationships. Let us take up these stages of analysis one by one.

### Selection of Units

Often the message materials to be analyzed will fall into natural units. One would normally take each day’s entry in a personal diary, for example, as a single unit. Or in analyzing the association structure of “Republicans” vs. “Democrats,” where a sample of individuals in each class have written letters to an editor, the letter from each individual would be a natural unit. Similarly, in studying the editorials in a certain newspaper, each editorial might be a unit. On the other hand, one may wish to analyze the contingencies in a more or less continuous message, for example in James Joyce’s *Ulysses*, and here it would be necessary to set up arbitrary units.

If the unit is too small (a single word, for example), then nothing can be shown to be contingent with anything else; if it is too large (the entire text or message, for example), then everything is completely contingent with everything else. There seems to be a broad range of tolerance between these limits within which approximately the same contingency values will be obtained. . . . In one small-scale investigation, . . . we found contingency values to be roughly constant between 120 and 210 words as units.

### Selection of Coding Categories

Here, as in most other types of content analysis, the nature, number, and breadth of

## 114 • PART 3

categories noted depend upon the purposes of the investigator. If the analyst has a very specific purpose, he will select his content categories around this core. In our own work, which has been methodologically oriented, we have merely taken those interesting contents most frequently referred to by the source. The same categorizing problems faced elsewhere are met here as well; for example, whether references to RELIGION in general, CHRISTIANITY, and the CHURCH should be lumped into a single category or kept separate. Of course, the finer the categories used, the larger must be the sample in order to get significant contingencies. We do run into one special categorizing problem with the contingency method, however: if one were to code two close synonyms like YOUNG WOMEN and GIRLS as separate categories, he would probably come to the surprising conclusion that these things are significantly dissociated in the thinking of the source; being semantic

alternatives, the source tends to use one in one location and the other in another location. If such closely synonymous alternates are treated as a single category, the problem does not arise.

### Raw Data Matrix

Armed with a list of the content categories . . . , the coder inspects each unit of the material and scores it in a raw data table such as that shown as Figure 1A.

Each row in the table represents a different unit (1, 2 . . . n) and each column a different content category (A, B . . . N). The coder may note merely the presence or absence of references to each content category; if present in unit 1, category A is scored plus, and if absent in unit 1, category A is scored minus—how often A is referred to (within a unit) is irrelevant in this case. One may also score in terms of each category being above or below its own median frequency; if above, plus, if

A. Raw Data Matrix					
Content Categories					
Units	A	B	C	...	N
1	+	-	+	etc.	
2	-	+	-		
3	-	+	-		
:	+	+	-		
n	etc.				
%	0.40	0.20	0.60		

B. Contingency Matrix					
Content Categories					
Units	A	B	C	...	N
A	-	0.08	0.24	etc.	
B	0.06	-	0.12		
C	0.38	0.02	-		
:	etc.				
N					

Obtained Contingencies  $P_{AB}$  →

Expected Contingencies  $P_A \times P_B$  ←

**Figure 1** (A) Raw Data Matrix and (B) Contingency Matrix

below, minus. This method needs to be used when units are relatively large and many categories tend to occur in most units (as can be seen, the presence/absence method in this case would show everything contingent on everything else). In this case, one first enters the actual frequencies of reference in the cells of Figure 3A, computes the median for each column, and then assigns each cell a plus or a minus depending on whether its frequency is above or below this median.

### Contingency Matrix

The contingency matrix, as illustrated in Figure 2B with entirely hypothetical data, provides the information necessary for comparing expected or chance going-togetherness of categories with actual obtained going-togetherness. The expected or chance contingency for each pair of columns is obtained by simply multiplying together the sheer rates of occurrence of these two categories, that is,  $p_A$  times  $p_B$  in analogy with the probability of obtaining both heads (HH) in tossing two unbiased coins whose  $P_H$  are both .50. We find the probabilities or relative rates of occurrence for each content category in the row labeled "percent" at the bottom of the raw data matrix. Thus, since A occurs in 40 per cent of the units and B in 20 per cent, we would expect A and B to occur together (be contingent) in only 8 per cent of the units on the basis of chance alone. Extending this to all possible pairs of categories, we fill in the upper right cells of the matrix, A/B, A/C, B/C, etc.

In the corresponding lower left cells of this matrix, for example B/A, C/A, C/B, etc., we then enter the actual or obtained contingencies; these are simply the percentages of units where plusses occur in both of the columns being tested. For example, in the part of the matrix shown in Figure 2A there is one such double plus between columns A and B.

If the obtained contingency is greater than the corresponding expected value (e.g., C/A .38, A/C .24), these events are co-occurring more often than by chance; if the obtained contingency is less than the corresponding

expected value (e.g., C/B .02, B/C .12), these events are co-occurring less often than by chance.

### Significance of Contingencies

The significance of the deviation of any obtained contingency from the expected value can be estimated in several ways. Baldwin (1942) utilized the chi-square test, in which a two-by-two frequency table (AB, A but not B, B but not A, neither A nor B) is arranged from the data in each pair of columns in the original data matrix and where the total N equals the number of units. This becomes pretty laborious with a large number of units. . . . [F]urthermore, the frequency of entries in the AB cell may often be below five, a number usually given as a lower limit in applying chi-square. We have used the simple standard error of a percentage,

$$\sigma_p = \sqrt{\frac{p(1-p)}{N}},$$

where  $p$  is the expected value in the upper right half of the contingency table and  $N$  is the total number of units sampled. This gives us an estimate of how much an obtained percentage may be anticipated to vary about its expected value; for example, if the sigma is .07 then a difference between the expected and obtained of .14 (two sigma) would only occur about five times in a hundred (two-tail test, direction of difference unspecified) by chance alone. . . . [W]ith large numbers of units, the size of  $p$  may become so small that some correction (e.g., an arc-sin transformation) must be made. . . . This method of estimating significance is not altogether satisfactory, and some work on a better method is needed. . . .

### Representation of Results

There are a number of ways in which the results of a contingency analysis can be represented, all of them being matters of convenience and efficiency in communicating



rather than rigorous quantitative procedures in themselves. (1) *Table of significant contingencies*. The simplest summary picture is a table which simply lists, for each category, the other categories with which it has significant associations or dissociations. (2) *Cluster analysis*. From the total contingency matrix, one may by inspection select sets of categories which form clusters by virtue of all having either significant plus relations with each other or at least include no significant minus relations. All such sets may be represented in an ordinary two-dimensional surface as overlapping regions (see Figure 2, the Goebbels diary data).

(3) *Models derived from the generalized distance formula*. Where the plusses and minuses in the raw data matrix represent frequencies above and below the median frequencies for each column, one may use the generalized distance formula

$$D = \sqrt{\sum d^2},$$

where  $d$  represents the difference in each unit between values (+ or -) in any two columns (zero where they have the same sign, 2 where different signs). If all signs between two columns are identical,  $D$  equals zero; if there is no correspondence,  $D$  is maximal.

We may now construct a new matrix similar to the contingency matrix (Fig. 1B) in which we enter  $D$  for every pair of categories. If no more than three factors are required to account for the relations in the  $D$  matrix, the entire set of distances can be represented in a solid (three-dimensional) model. If more factors are involved, a three-dimensional (representation) can only approximate the true distance relations (even though the values in the  $D$  matrix are valid for any number of dimensions) (see Osgood & Suci, 1952). (The reason the  $D$  method cannot be applied where mere presence and absence is recorded is that in this case pairing of minuses between

columns merely indicates lack of relation or independence between categories.)

#### ILLUSTRATIVE APPLICATIONS OF THE CONTINGENCY METHOD

##### Cameron's Ford Sunday Evening Hour Talks

A sample of 38 talks given by W. J. Cameron on the Ford Sunday Evening Hour radio program, each talk running to about 1,000 words, was studied by this method.<sup>3</sup> Each talk was treated as a unit. Based on a preliminary reading, 27 broadly defined content categories were selected in terms of frequency of usage. The analyst then went through these materials noting each reference to these categories. The median frequency of appearance of each category was computed and a matrix of units (rows) against categories (columns) was filled. A plus sign was entered for frequency of category above the median for a unit and a minus for frequencies below. Applying the formula for  $D$  given above to each pair of content columns, a  $D$  matrix showing the distance of each category from every other category was computed.\*

... [Those] familiar with Cameron's talks [have acknowledged] that the pattern of relationships produced here had considerable face validity. References to FACTORIES, industry, machines, production, and the like (FAC) tended to cluster with references to PROGRESS (PRO), FORD and Ford cars (FD), free ENTERPRISE and initiative (ENT), BUSINESS, selling, and the like (BUS), and to some extent with references to RUGGED INDIVIDUALISM, independence (RI), and to LAYMEN, farmers, shopkeepers, and so on (LAY). But when Cameron talked about these things he tended *not* to talk about (i.e., to dissociate them from) categories like YOUTH, our young people (YTH), INTELLECTUALS, "lily-livered bookmen," etc.

\*This can be represented in three dimensions; see Krippendorff, *Content Analysis*, 207.

(INT), and DISEASE, poisoned minds, unhealthy thoughts, and the like (DIS), which form another cluster. This relation in Cameron's thinking between YOUTH (always favorable) and DISEASE notions (obviously unfavorable) was unsuspected by the analyst until it appeared in the contingency data, which suggests one of the potential values of the method.

Also tending to be dissociated from the FORD, FACTORIES, ENTERPRISE cluster, and more or less independent of the YOUTH, INTELLECTUALS, and DISEASE one, we find an interesting collection of superficially contrary notions: on one hand we have SOCIETY in abstract, civilization (SOC), CHRISTIAN, God, and church (CHR), our ELDERS, mature minds (ELD), TRADITION and basic values (TRAD), and to some extent the PAST of our forefathers (PAS) and our HOMES, fireside, and families (HOM)—all things favorably drawn; but on the other hand, in the same cluster, we find DESTRUCTION and violence (DES), assorted ISMS like Communism, Fascism, and totalitarianism (ISM), FEAR, bewilderment, and dismay (FEAR), and sundry EVILS (EVL). Apparently, when he thinks and writes about the solid, traditional things that hold society together, he immediately tends to associate them with the things he fears, the various isms that threaten destruction of his values. References to the FUTURE (FUT) and to our HOPES and confidence in the New World (HOP) tend to be associated with references to AMERICA (AM), but also ISMS again. The allocation of a few other notions, including references to the general PUBLIC (PUB), to FREEDOM and democracy (FREE), and human NATURE, what is instinctive or natural (NAT), may be studied by the reader himself.

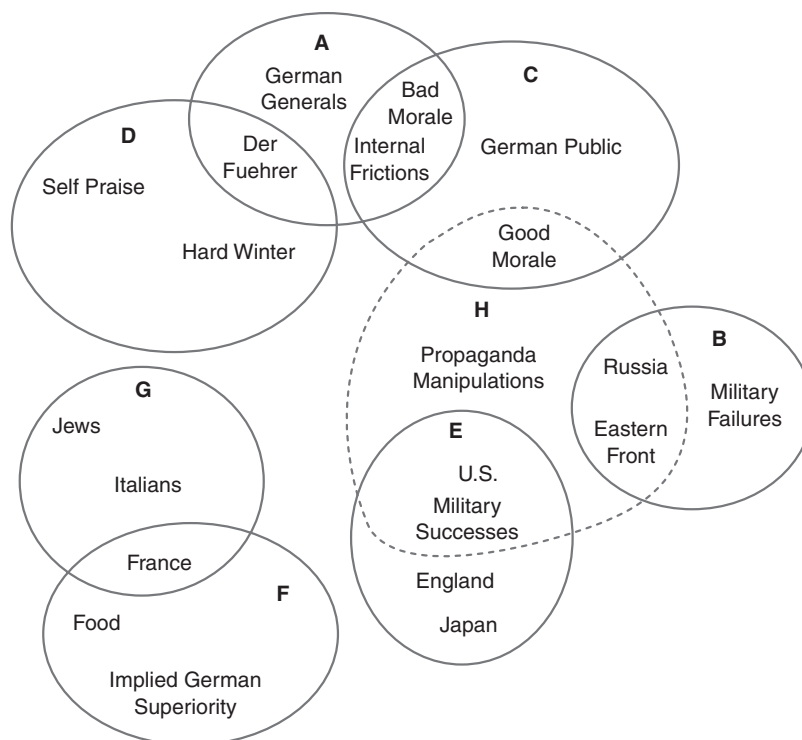
### Goebbels' Diary

Using a table of random numbers to select pages and then lines-on-page, 100 samples, each approximately 100 words in length (beginning and ending with the nearest full

sentence), were extracted from the English version of Goebbels' diary and typed on cards. An example would be:

#38. Spieler sent me a letter from occupied France. He complained bitterly about the provocative attitude of the French, who continue to live exactly as in peacetime and have everything in the way of food that their hearts desire. Even though this is true only of the plutocratic circles, it nevertheless angers our soldiers, who have but meager rations. We Germans are too good-natured in every respect. We don't yet know how to behave like a victorious people. We have no real tradition. On this we must catch up in the coming decades.

In terms of a rough frequency-of-usage analysis made previously, 21 content categories were selected for analysis. An independent coder went through the 100 units in a shuffled order noting simply the presence or absence of reference to these 21 categories, generating a raw data table like that illustrated in Figure 1A. The data were then transformed into a contingency table of the sort shown as Figure 1B, and significance tests were run (utilizing the arc-sin transformation). References to GERMAN GENERALS were significantly contingent upon references to INTERNAL FRICTIONS (in the inner circle about Hitler) at the one per cent level; references to GERMAN PUBLIC were associated with those to BAD MORALE at the 5 per cent level, as were contingencies between RUSSIA and EASTERN FRONT; negative contingencies, significant at the 5 per cent level, were obtained between RUSSIA and BAD MORALE, between references to ENGLAND and references to GERMAN SUPERIORITY as a race, and between references to the GERMAN PUBLIC and references to RUSSIA. Such negative contingencies are at least suggestive of repressions on Goebbels' part; that is, avoiding thinking of Russia when he thinks of the bad home-morale situation, avoiding thinking about England when he thinks about the superiority of the German race, and so on. These are merely inferences, of course.



**Figure 2** Clusters of Contingencies; Goebbel's Diary

A cluster analysis was made of these data, with the results shown in Figure 2. The content categories included within regions have mainly plus relations and no minus relations. Numerous inferences might be made from this chart. For example: (D) that Goebbels defends himself from thoughts about the HARD WINTER with SELF PRAISE and thoughts about his closeness to DER FUEHRER; (A) that ideas about BAD MORALE lead promptly to rationalizations in terms of the INTERNAL FRICTIONS brought about by GERMAN GENERALS, which in turn bring up conflicts between himself and others in securing the favor of DER FUEHRER; (C) that thoughts about his job of maintaining GOOD MORALE among the GERMAN PUBLIC lead to thoughts about BAD MORALE and INTERNAL FRICTIONS; (H) that his problem-solving ideas about PROPAGANDA

MANIPULATIONS may lead him alternatively to the GOOD MORALE cluster of associations, to the dismal RUSSIA-EASTERN FRONT-MILITARY FAILURES cluster, or to the more encouraging cluster in which his ally, JAPAN, is having MILITARY SUCCESSES against ENGLAND and the U.S.; and finally (G and F), that when he thinks about the subject peoples, JEWS and ITALIANS, and FRANCE, he tends also, particularly in the case of FRANCE, to think about difficulties of maintaining FOOD supplies, leading quite naturally to ideas about GERMAN SUPERIORITY in withstanding hardships, and the like. These are inferences, of course; there are alternative interpretations possible as to why any cluster of symbols shows positive or negative contingency. But the inferences have the advantage of resting on demonstrable verbal behavior, which may even be unconscious to

the source. They do not necessarily depend upon explicit statements of relation by the source.

#### CRITIQUE OF THE CONTINGENCY METHOD

The use of the contingency method is based upon a very general inference between messages and those who exchange them. . . . [C]ontingencies . . . in messages are indicative of the association structure in the source and predictive of the association structure that may result in the receiver (given sufficient frequency of [exposure]). But under what conditions is this general inference . . . valid?

If we are dealing with spontaneous informal messages from a single known source (e.g., personal diaries, . . . letters to friends and family, . . . extemporaneous speech, as in psychotherapeutic interviews, etc.) then attribution of the association structure to this source is probably most defensible. When dealing with deliberately planned messages, particularly when the source is an institution, as . . . in propaganda (or mass media content) analysis, it would probably be safer to speak of the “policy” of the source rather than its association structure.

(What does the contingency method yield when language use is instrumental or cynical, as can be expected when the analyst faces a clever rhetorician, a propagandist, an advertiser, a client in therapy, or a candidate for a political office? Those who question the use of the contingency method under such conditions assume that the results must represent message contents. This is not so.) The fact that references to YOUTH and DISEASE by Cameron are significantly associated says nothing (about Cameron’s belief system) about the direction of the assertions between them. . . . Cameron’s typical statement would

be that “Our young people are *not* susceptible to the diseased ideologies of our times.” What the method tells us, however, is that there is a greater-than-chance tendency for ideas about DISEASE to occur in the environment of ideas about YOUTH—quite apart from what assertions he may make relating these two.

[We] assume that a significant contingency, whether positive or negative, is evidence for an underlying association (not for whether they are habitual or deliberate). If the contingency is negative (i.e., a significant dissociation) it presumably means that these ideas are associated with some kind of unpleasant affect. (Intentionally avoiding certain associations, for example, in order not to offend somebody, to hide something, or in compliance with a taboo, suggests that the concepts are close in the mind of the source—not much else.)

The contingency method . . . does not take into account the (expressed) intensity with which assertions are made—if the source says “The French are *definitively* like the Italians in this respect,” the method only records an instance of contingency between the categories of French and Italian. On the other hand, reflecting the basic psychological principle relating habit strength to frequency of response, the method does indirectly reflect the strength of an association (or dissociation).

. . . [A]ssociation is not indicative of semantic similarity. References to COMMUNISM may frequently lead to reference to CAPITALISM, but this does not necessarily imply that these concepts are either similar in reference or in psychological meaning. (They exhibit a contrast within a common linguistic domain, as in GOD and DEVIL, SOLDIER and SAILOR (or BUYER and SELLER).\*

\*Another example for the fact that associations have little to do with semantic similarities is synonyms. Synonyms rarely co-occur near each other, and lacking contingency might give the impression of dissociation. In contingency analysis, as presented here, this problem does not arise as it is applied to categories that subsume synonyms. When applied to raw words (see reading 7.3, this volume) this becomes a distraction.

## NOTES

1. This experiment was done by the author in collaboration with Mrs. Lois Anderson (1957).

2. These tables and certain details of statistical treatment will be found in Osgood and Anderson (1957).

3. It is instructive to compare the results of this contingency analysis with an earlier frequency analysis of speeches by the same speaker by Green (1939). The clusters spotted in the present study seem to have been largely overlooked in Green's more conventional analysis. The studies were independent of each other. The earlier one was not known to the present author.

## REFERENCES

- Baldwin, A. L. (1942). Personal structure analysis: A statistical method for investigating the single personality. *Journal of Abnormal and Social Psychology* 37:163–183.
- Green, Jr., T. S. (1939). Mr. Cameron and the Ford House. *Public Opinion Quarterly* 3:669–675.
- Osgood, C. E., & Anderson, L. (1957). Certain relations between experienced contingencies, association structure, and contingencies in encoded messages. *American Journal of Psychology* 70:411–420.
- Osgood, C. E., & Suci, G. J. (1952). A measure of relation determined by both mean difference and profile information. *Psychological Bulletin* 49:251–262.

## 3.2

# FOUR TYPES OF INFERENCE FROM DOCUMENTS TO EVENTS

VERNON K. DIBBLE\*

Many of the intellectual procedures used by historians can be viewed in terms of the dichotomy between documentary evidence and facts or events that are external to the documents themselves. At some moments, historians work on only one side of this divide. Where the meaning of a document is not clear, for example, they sometimes use one phrase in order to infer the meaning of another. Or, working only on the other side of this dichotomy, once a given fact is established, they use it in order to make inferences about other facts. Although such inferences are almost completely neglected by manuals of historical method, they are very common in the works of historians. Chrimes (1952:15–16) writes, for example, that “from the time of Cnut at least we begin to see men who began their careers as Scribes in the king’s service blossoming forth to be bishops and abbots—a sure sign of their growing importance and favor.” The rules for making inferences of this type, whatever they might turn out to be when adequately codified, are quite different from rules for the use of

documents. Once the first fact is established (e.g., the career lines of royal scribes), the historian’s inference to the second fact (e.g., the importance of the royal secretariat) has nothing to do with documentary techniques.

At other moments historians move from one side of the dichotomy to the other. Moving from fact to document, they often use the former in order to make inferences about the provenance, age, authenticity, or authorship of the latter. And, of course, they also move in the opposite direction, using documents in order to make inferences about external events.

Although these various procedures are used in conjunction with one another, and although the resulting conclusions often stand or fall together, this article is concerned only with inferences from *documents to events*. It identifies four quite different ways in which historians make such inferences, as illustrated in recent historical literature, and discusses some of the problems which each of the four entails. Manuals of historical method also concentrate on inferences from documents to

\*From Dibble, V. K. (1963). Four types of inference from documents to events. *History and Theory* 3:203–221.

events. But they hardly reflect the procedures that historians actually use. To judge by most manuals, historians establish facts from documents primarily by examining *testimony* to events, which is recorded by witnesses who have seen or heard about these events. This article is concerned not only with testimony but also with three other categories, which may be termed *social bookkeeping*, *correlates*, and *direct indicators*.

These four categories are based on two very different criteria of classification. First, the distinction between testimony and social bookkeeping is a classification of sources. Of all documents used by historians, some purport to record information about things that happen, and some do not. Codes of law, pieces of pottery, and poems, for example, do not. Documents, which do purport to record information, can in general be classified further as testimony or social bookkeeping, depending upon the circumstances under which they are produced. Since the procedures appropriate to testimony are not identical with those appropriate to social bookkeeping, this classification of sources is simultaneously a classification of techniques. Second, the distinction between correlates and direct indicators refers only to techniques. For all documents, both those that purport to record information and those that do not, are potential correlates and potential indicators. This double-edged scheme of classification must be kept in mind as we proceed further.

### TESTIMONY

The manuals give a number of familiar rules for evaluating testimony. Many historians do not care for generalizations or formal methodology and prefer to regard masterful documentary criticism as “a sort of sixth sense that will alert you to the tell-tale signs.”<sup>1</sup> But most of these rules can be stated as general laws in one or another social science, although they are all probability laws and many are definitely of the armchair variety. Some rules turn out to be general laws governing the psychology of cognition: testimony about specific details is likely to be more accurate than testimony about

general conditions.<sup>2</sup> Others are laws governing the psychology of memory: testimony recorded shortly after an event took place is likely to be more accurate than testimony recorded long afterwards. Other rules can be stated as general laws, which govern communication: testimony about ideologically relevant events, which is addressed to people who share the witness’s beliefs and values, is likely to be more accurate than testimony addressed to audiences that do not share the witness’s ideology. Some rules turn out to be laws governing cultural processes in cognition: the rule of thumb that the ancients grossly overestimated numbers would be such a law, if we were able to state in what kinds of societies or cultures people overestimate numbers and in what kinds they underestimate them.

In using such rules, historians implicitly construct syllogisms that include probability statements. That is, they begin with premises, which are stated in terms of likelihood rather than certainty and, therefore, proceed to conclusions, which are likely rather than certain to be true. This logical structure of inferences from testimony to events is seen more clearly when historians choose between *n* conflicting accounts than when they evaluate a single piece of testimony. . . .

Historians reach an overall conclusion at such points by engaging in a peculiar kind of arithmetic without numbers. They assign weights to each syllogism and to its conclusion, usually assigning greater weight to some than to others. . . . The different weights, or estimates of probable accuracy, are compared with one another, and out of the probabilities ascribed to each syllogism, an aggregate probability for each of the possible overall conclusions is arrived at. These combinations of probabilities, and comparisons between them, are carried out despite the fact that they are never stated with quantitative precision. For the logic of inferences from testimony to events is the logic of qualitative probabilities.

But while the notion of probability or likelihood may apply to the historian’s assessments and conclusions, it does not apply to the single event in question. Patrick Henry either did or did not profess his loyalty to the king, and there is no probability about it. To say

historians infer from documents to events by the logic of qualitative probabilities is to say that if they make inferences about hundreds of events by the simultaneous application, to each event in question, of a number of syllogisms which include probability statements, then they will reach the correct conclusion more often than not. But since their premises are probability laws, the most rigorous evaluation of testimony can lead to the incorrect conclusion in any single case. If, as the manuals suggest, historians relied primarily on witnesses, if document equaled testimony and internal criticism equaled the evaluation of testimony—then historians would seldom be able to spot those instances where rigorous evaluation of testimony leads to the wrong inference. But the impression conveyed by the manuals is fortunately incorrect. Historians infer from documents to events in ways that have nothing to do with the evaluation of testimony. Other types of inferences are sometimes used without reference to testimony and sometimes supplement testimony. The first of the three remaining types to be considered here is the use of social bookkeeping.

#### SOCIAL BOOKKEEPING

When the manuals speak of witnesses and their testimony and of the rules for evaluating testimony, there is always the implicit assumption that documents are produced by individuals and not by social systems. The secluded monk, the diarist alone in his room at the end of the day, and the solitary traveler are the classic examples of the historian's witness. But if one were to enumerate the sources used by, say, fifty representative historians, then one would probably find that documents produced in individualized circumstances make up only a small percentage of the total and are outnumbered by social bookkeeping. Groups and organizations in all literate societies have institutionalized procedures for recording facts and events. The term *social bookkeeping* refers to all documents which purport to record information and which are the product of groups and organizations. The term includes such diverse sources as

transcripts of parliamentary debates, calendars of saints, bankbooks, tax returns, inventories of estates, the Domesday Book, court records, crime statistics, censuses, reports by subordinates in hierarchies to their superiors, and the list of graduates of Harvard University.

Testimony is the work of individuals. Historians have accordingly evolved a psychology and social psychology of documents that guide them in their use of testimony. Social bookkeeping is the work of social systems. But historians have not yet evolved a sociology of documents to guide them in their use of such sources. One does find in the manuals a few stray reminders that documents of this type must be read in the light of the social system that produces them. Students are reminded, for example, that the *Congressional Record* is not a literal transcript of Congressional debates. The *Record* is an inaccurate transcript not because recording clerks have faulty hearing, or political biases, or any other failing to which witnesses are prone. It is an inaccurate transcript because of one simple feature of the social system that produces it: members of Congress are free to amend their remarks before the *Record* goes to press.

It is possible to state a few general principles for the use of social bookkeeping. The *Congressional Record* reminds us that different forms of social bookkeeping vary in the extent to which interested parties have a hand in producing the record. In some societies, inventories of estates are compiled by heirs and in others by disinterested parties. The *Record* also reminds us that different forms of social bookkeeping vary in the extent to which interested parties are likely to check the record after it is first set down. People are more likely to check certificates of inheritance or deeds to their land than the information about themselves, which is collected by census enumerators. Different forms of social bookkeeping, which are checked by interested parties, vary in the extent to which the interested parties are free to alter the record. Lords of manors could presumably alter records, which were compiled by their own overseers, more easily than they could alter records compiled by tax officers. Different



forms of social bookkeeping, which can be altered by interested parties, vary in the extent to which such alteration makes for greater accuracy or for less. Transcripts of some legal proceedings are less complete than they would be otherwise because interested parties can sometimes have remarks “stricken from the record.” In contrast, alterations initiated by an interested party to his own advantage are likely to make for a more complete and a more accurate record if the record-keeper is in a position to make his own independent check on the accuracy of the suggested alteration. The professor who asks his chair(person) or dean to add missing items to his bibliography is an obvious example. If the record-keeper is not in a position to make an independent check, as with compilations of research allegedly in progress, the interested party is free to embellish veracity.

Some forms of social bookkeeping are provided with built-in checks, apart from interested parties, while others are not. The Bollandist fathers provide the Catholic Church with an institutionalized check on the record of saints: there is no built-in check on the biographies submitted to the editors of *Who's Who*. Different forms of social bookkeeping vary in the extent to which the events recorded are visible to the record-keeper or in the extent to which communication between observers and record-keepers is assured. They also differ in the number of steps between observer and record-keeper. There are no steps between observer and record-keeper when, as with court stenographers, the two jobs are performed by the same person. There are many steps in the hierarchy of a corporation between a sales(person)'s weekly reports to an immediate supervisor and the record of sales in the corporation's annual report, with communication steps in the sales department paralleled by different steps in accounting and billing departments. Different examples of social bookkeeping, which do come into being only after many steps between observer and record-keeper, vary in the extent to which distortion or suppression of information takes place along the way. Staff officers in contemporary American corporations get line personnel to innovate by agreeing to distort budget

reports in order to make line personnel look better than they really are; no matter how many steps there might be between graduate students, dissertation supervisors, departmental secretaries, deans' offices, and printers of commencement programs, carelessness is not the only thing which might distort or suppress information along the way.

Some of these general observations, or others like them, have been concretely applied in the works of historians who have had to come to terms with particular kinds of social bookkeeping in particular historical societies . . . Kosminsky's criticism of the survey of 1279 and of certain other forms of social bookkeeping in thirteenth-century England is an example (Kosminsky, 1956). He asks who initiated the survey, and why. To whom would the returns be valuable and who might be hurt by them? Who carried it out, and how? What did the officers of the king do when they arrived in a county? How did the local juries acquire their information? Did the royal officers check on the local juries or simply accept their returns as given? Was new machinery devised for gathering the information required for the surveyor or established and tried machinery used? In what respects did the questions presented to the local juries force them to simplify the facts? What was most visible to the juries and what was least visible?

Many of the questions posed by Kosminsky have exact parallels in the criticism of testimony. Comparisons between descriptions of a manor in the survey of 1279 and in an *Inquisition post mortem* are analogous to comparisons between the testimonies of two witnesses to the same event. In some cases, however, we do not really have two independent records, since the information given by local juries was sometimes copied from another source. This is, of course, parallel to the difference between two independent witnesses and two witnesses, one of whom reports what the other had told him. Kosminsky's examinations of the vocabulary of the survey and of its internal consistency also have their parallels in the evaluation of testimony. For other questions, however, there is no parallel with the evaluation of testimony. The difference between improvised

and established machinery for acquiring information, questions concerning the flow of communications and commands between different people involved in gathering the data, and questions concerning the extent to which some people checked up on other people, all point to the distinctly social character of social bookkeeping. In such questions as these the historian is concerned not with the veracity or eyesight of individuals, but with the operation of social systems.

As with criticism of testimony, historians criticize social bookkeeping in order to make decisions about the probable accuracy or completeness of the record. But, of course, historians are not interested only in accurate social bookkeeping. Inaccurate social bookkeeping can be just as valuable as testimony known to consist of lies and distortions. Whatever the survey of 1279 tells us or fails to tell us about manors and villages, it also tells us something about the administrative mechanisms of the medieval English state. Documents, which purport to record information, to be useful to historians, need to be accurate only when historians are concerned with the information they purport to record. But testimony is not always used qua testimony and social bookkeeping is not always used qua social bookkeeping. Documents of both types are used as correlates or direct indicators of facts or events other than those they purport to record. And, of course, documents that do not purport to record information can be used in the same way.

#### DOCUMENTS AS CORRELATES

Historians are often able to make inferences from documents to events in the absence of testimony or social bookkeeping that tells them about the events in question. One way of doing so is to use documents whose characteristics are known to be correlated with the events in question. Haskins (1918) provides a particularly striking example of the use of documents as correlates, supplementing the use of testimony.

In *Norman Institutions*, Haskins demonstrates that the governmental machinery of Robert Curthose (1087–1096 and 1100–1106)

was weak, ineffective, and underdeveloped, but that the more highly developed institutions of the Conqueror were “in some measure maintained even during the disorder and weakness of Robert’s time” (Haskins, 1918:84). Among his sources are the narratives of Odericus Vitalis, the charters of Robert Curthose, and the charters of William Rufus, who ruled Normandy between 1096 and 1100 while his brother Robert was on a crusade. The narratives of Odericus are, of course, an example of testimony. His descriptions of Normandy under Robert are “a dreary tale of private war, murder, and pillage, of perjury, disloyalty, and revolt . . .” (Haskins, 1918:62). Of William Rufus, in contrast, “Odericus tells us that . . . under his iron heel Normandy at least enjoyed a brief period of order and justice to which it looked back with longing after Robert’s return” (Haskins, 1918:80).

Haskins is less concerned with private war and public peace than with the institutions of Norman government. The testimony of Odericus is not adequate for his purpose, since the witness was not close to Robert’s governmental machinery and since his perceptions were colored by his geographical location and by his position as a monk. Haskins makes inferences about organs of government from the testimony of Odericus and from certain other narratives but then gives reasons not to rely on these inferences:

Amidst these narratives of confusion and revolt, there is small place for the machinery of government, and we are not surprised that the chroniclers are almost silent on the subject. Robert’s reliance on mercenaries [reference to Odericus and to another witness] shows the breakdown of the feudal service, which may also be illustrated by an apparent example of popular levies [reference to Odericus]; his constant financial necessities [reference to Odericus and to another witness] point to the demoralization of the revenue. The rare mention of his curia [reference to Odericus] implies that it met but rarely. Still, these inferences are negative and to that extent inconclusive, and even the detailed account of Odericus is largely local and episodic, being chiefly devoted to events in the notoriously troubled region of the south, and is also colored by the sufferings and losses of the church. (Haskins, 1918:64)

Thus far, Haskins could be following the injunctions of the manuals concerning the use of testimony. His next step, however, has not been dreamt of in the manuals. Having found his star witness wanting, he turns to the charters and similar documents of Robert's reign and uses certain of their characteristics as correlates of the nature of Robert's government. He notes that the number of surviving charters is small, relative to the length of the reign and in comparison with other Norman dukes. Perhaps only thirty-nine survive because "later times were indifferent to preserving charters of Robert Curthose, but it is even more likely that his own age was not eager to secure them. As confirmation at his hands counted for little, none of these charters consist of general liberties or comprehensive enumerations of past grants; they are all specific and immediate. Furthermore, so far as can now be seen, the surviving documents are all authentic; privileges of the Conqueror, Henry I, or Henry II were worth fabricating but no one seems to have thought it worth while to invent a charter of Robert" (Haskins, 1918:71).<sup>3</sup> Seventeen of the existing charters were issued in Robert's name, while twenty-two were drawn up by interested parties for him to attest. The seventeen issued in Robert's name are not uniform in size, style, or method of authentication. Of the seven, which are preserved in the original, each is in a different handwriting. Seals were used on only some of the charters, and were not used in uniform fashion. There are nine variations on the title *dux Normannorum*. To these varying titles there is sometimes added one of three variations of *filius Willelmi gloriosi regis Anglorum*. Robert signs sometimes as *dux* and sometimes as *comes*. Some charters invoke the Trinity while others do not.

From these and similar observations, Haskins concludes "the range of variation in style and form precludes the existence of an effective chancery and indicates that the duke's charters were ordinarily drawn up by the recipients" (Haskins, 1918:74). The decline of the ducal chancery is accompanied by a decline in the *curia*. The lists of witnesses on charters show little continuity in the ducal entourage and "still less any clearly

marked official element" (Haskins, 1918:76), and a meeting of the *curia* is mentioned only once in the surviving documents. In short, the characteristics of Robert's charters confirm the inferences about the nature of his rule made on the basis of testimony in the chronicles. A similar examination of the writs and charters issued in or about Normandy by William Rufus during his reign there also confirms the testimony of Odericus. Under William Rufus, Haskins infers, we see "the regular mechanism of Anglo-Norman administration at work" (Haskins, 1918:83).

Haskins's use of ducal charters has nothing to do with the evaluation of testimony. And although charters are a form of social bookkeeping, since they record information about grants and privileges given by the crown, Haskins is not using them as such. He is not primarily concerned with the information about grants and privileges they contain. Haskins's procedure illustrates the making of inferences from documents to events by the use of documents as correlates of the events in question. As with testimony and social bookkeeping, the logic of such inferences can be stated syllogistically: Norman dukes known to have effective organs of government issued charters with the characteristics a, b, and c; Duke Robert's charters have characteristics which are the opposite of a, b, and c; therefore, Robert's rule must have been weak.

There are, of course, a number of syllogisms here, one for each characteristic in question. And, as with testimony and social bookkeeping, historians somehow combine the differently weighted conclusions of each syllogism in order to reach an overall conclusion, even though the weights to be given to each conclusion are not precisely known. It should be noted further that Haskins's major premises are not strictly adequate to the conclusion. Ideally, he should have grounds for his major premise that Norman dukes known to have effective organs of government issued charters with given characteristics while Norman dukes known to have ineffective organs issued charters with the opposite characteristics. Haskins actually has grounds only for making the first part of the statement, simply because there were not enough weak

dukes to provide the evidence for the second part. The problem is hardly serious, however. Our knowledge of Norman government and society, and of the functions of charters in such a system, allows us to state why the characteristics of ducal charters should be correlated with the strength of ducal governments.

It may some day be possible to state general rules for the use of documents as correlates, just as there are rules in the manuals for the use of testimony. But most of these rules are likely to be quite different from the rules concerning testimony. The latter are, in effect, psychological or social psychological laws governing such phenomena as cognition, memory, and communication. Although a few similarly general principles might one day be stated for the use of documents as correlates, it is usually possible to use documents in this way only because of historically specific knowledge about the institutions of particular societies or types of societies. It was such knowledge, and not general laws, which enabled Haskins to use charters as correlates of the nature of Robert's regime.

#### DOCUMENTS AS DIRECT INDICATORS

The fourth type of inference from document to event might appear to entail no inference at all. This is the case when all or part of the document itself, as opposed to external events which are recorded by or correlated with the document, is the datum under investigation. If one wants to know, say, what the British ambassador in Berlin reported to the Foreign Office on the day Bismarck moved against France, then one simply finds and reads whatever messages he sent. His cables or dispatches are direct indicators. What is more, if no records have been lost and if there were no oral messages that were never recorded, then the documents themselves provide an exhaustive answer to the historian's question. There is no need to infer from documents to events. Direct indicators, surely, have nothing to do with inferences.

The matter is not quite so simple, however. The example given illustrates the methodologically uninteresting case in which the content

of the documents and the answer to the historian's question are completely coterminous with each other. Documentary research comes closer to absolute certainty in such cases than in any other. There is no need for probabilistic syllogisms, which might lead to incorrect conclusions even when most rigorously applied. But this certainty is possible only in special cases, and sometimes requires that only trivial questions be asked. The documents at hand and the answer to the historian's question are coterminous only when two conditions are met. First, it must be possible to answer the question by reference to the documents themselves, as the historian's subjects happened to produce them, and without reference to their accuracy concerning, or correlation with, events external to the documents. Two sorts of questions meet this condition. (1) To continue with the example of Norman charters, one might ask questions about the formal characteristics of documents. What were Robert Curthose's charters like? (2) One might ask questions, which, in effect, simply state the content of the documents in interrogative form. What rights or privileges were granted to what monasteries in which of Robert's charters? Since a ducal grant of privileges in Norman society may be defined as the emission or attestation by the duke of a charter which states that he is making such a grant, charters can be used as direct indicators of the events in question. There is no need to argue the contention that historians cannot limit themselves to questions of these two types.

Second, even though the historian might be posing questions of these two types, the documents at hand and the answers to his questions are rarely if ever coterminous unless the questions call for purely descriptive answers. Such answers are not sufficient when historians conceptualize. And, as Marc Bloch has taught us, historians conceptualize all the time, even when they deny all interest in concepts and claim to deal only with unique particulars. A glance at one example of conceptualization in the work of a historian will indicate why conceptualization usually makes it impossible to pose questions whose answers are coterminous with the documents themselves.

As everyone knows, to conceptualize is to classify a number of relatively less general items under some relatively more general rubric. Instead of limiting themselves to specific items, on the one hand, and a single general rubric, on the other, historians and other social scientists often find it useful to specify conceptual rubrics of intermediate generality. Heckscher, for example, defines “mercantilist economic policy” in terms of five intermediate rubrics; mercantilism as an agent of unification, as a system of power, as a protectionist system, as a monetary system, and as a conception of society (Heckscher, 1955). Each of these intermediate rubrics or “five aspects of mercantilism,” not to speak of the more general rubric under which they are all subsumed, sums up an enormous variety of concrete details. Each of them refers to a vast number of books or treatises, acts of parliament, royal decrees, instructions sent down through administrative hierarchies and arguments over economic policy. If the problem were one of economic practice, such sources would not necessarily tell us what people actually did and it would be necessary to make inferences from documents to events external to the documents. But since policy is defined by what people say and not only by what people do, such sources may in most cases be taken as direct indicators of mercantilist economic policy.

Implicit in Heckscher’s definition of mercantilist policy is the empirical assertion that these “five aspects of mercantilism” were in fact correlated with one another, either because they were effects of the same cause or because some of them were causes of others. Also implicit in Heckscher’s definition, however, is the empirical assertion that the correlation between the five aspects of mercantilist policy is less than perfect and is not always observed. Although the first two components of mercantilism, “unification and power, were well suited to each other . . . it is . . . important to draw attention to the opposite point, that the two were not inseparable. That there were two separate aspects becomes clear in considering *laissez-faire*, for this policy usually combined a unification which was almost complete in every respect with a remarkable indifference to considerations of power” (Heckscher, 1955:24).

This formulation betrays a notion which Heckscher did not spell out explicitly: the five components of his definition could all be subsumed under the same concept because they tended to go along together empirically, but there were five “separate aspects” because they did not necessarily go along together completely or in all times and places. And if the correlation between these five general rubrics is less than perfect, what are we to think of the enormous number of details which each of these rubrics sums up? If we look at the details under any single rubric, under monetary policy let us say, would we expect to find that a highly mercantilist royal decree is necessarily followed by other decrees which are equally mercantilist? And that as royal decrees become more mercantilist all judicial decisions, opinions of high officials, books published, and instructions sent down to subordinates by administrative superiors will follow along? Of course not. There is more free floating among the details, among the specific indicators of the general concept, than is seen when we deal with conceptual classifications that are established by adding up the general tendency of the details.<sup>4</sup>

The lesson is clear. If historians want to use direct indicators in order to answer limited factual questions, which call for purely descriptive answers, then there is no problem. The ambassador’s cables are direct indicators of what he told the Foreign Office. But when historians use concepts that sum up a large number of details, the answers to their questions can rarely if ever be coterminous with any single document or set of documents. For any single item may not be a reliable indicator of a concept, depending upon the way in which such items are inter-correlated with other items that are also subsumed under the same concept. A mercantilist treatise published in France in 1690 is sufficient to indicate that France was a mercantilist nation, only if we know that the characteristics of all books on economic policy were highly correlated and that the characteristics of such books were in turn correlated with the characteristics of royal decrees, memoranda by high officials, and all the rest. When information on all items subsumed under the concept is

available, or when adequate samples are possible, there is no particular difficulty. When neither is possible, the problem may be handled in one of two ways. First, the indicators available might include such a large portion of all items included under the concept that we need not worry about the missing indicators. Whatever they might look like, if ever located, they could not radically change the original judgment. Second, even though the available indicators be a small portion of all items included under the concept, the historian might know or might be able to infer the ways in which they must have been inter-correlated with the missing items. The difference between these two situations is nicely illustrated in one of the most imaginative examples of the use of direct indicators that can be found in historical literature, V. H. Galbraith's guide to the Public Record Office in London (Galbraith, 1934).

Galbraith's task is to find order in collections that fell into disorder during centuries of neglect. In order to do so, he argues, one must regard documents as secretions of the organizations that issued them. Hence, he must reconstruct the issuing organizations and the relationships between them. His reconstruction of the various organs of government in medieval England illustrates the second situation described above. With hardly a glance at the substantive content of any document, he uses differences in seals, parchment, filing practices, systems of dating, language, and handwriting as indicators of the existence of separate organs of government and of their emergence at various points in time. Now, when one speaks of separate organs of government, and of the boundaries and differences between them, one refers primarily to patterns of interaction between officials or clerks and to the actions they perform. Characteristics of pieces of paper or of the way in which they are filed are only a small part of what one means. But the indicators that Galbraith uses must have been so highly correlated with the social systems within which officials and scribes did their jobs that we do not worry if they are only a small part of everything that is meant by "distinct organ of government" or "separate organization."

Galbraith's reconstruction of the changing channels of communication between the various organs of government in medieval England illustrates the first situation described above. Among fourteenth- and fifteenth-century documents, for example, he finds warrants under the Royal Signet, which were filed among the records of the Privy Seal office, and warrants under the Privy Seal, which were filed among the Chancery records. These documents indicate a flow of communication from the King through the office of the Privy Seal to the Chancery, a more complex system than is seen in earlier records. The documents indicate a similar flow of communication from the Council and other departments through the office of the Privy Seal and then on to the Chancery. The office of the Privy Seal was a great clearing house, the center of a system that placed the Great Seal, in the custody of the Chancery, at the disposal of all departments. But certain notations on Chancery documents, such as *per ipsum regem* or *per consilium*, indicate that these normal channels were sometimes circumvented. The King or other officials sometimes communicated directly with the Chancery. Communication between departments is indicated in other ways as well. The Exchequer's *Originalia* Rolls are extracts of Chancery Rolls dealing with fines payable, and duplicate copies of the Chancery's Inquisitions post mortem are found in the Exchequer archives.

"Channels of communication" includes oral communications which were never recorded and perhaps the less important, illicit, or informal communications which were not preserved. Some traces of oral communications are found in the documents, and written communications not destined for preservation are less of a problem when scarce parchment rather than plentiful paper was used. For these reasons, we need not worry too much about the missing indicators. Those, which are available to Galbraith, cover such a large portion of the phenomena in question that additional evidence could not change his judgments by very much. To speak of cooperation between departments is quite another matter (Galbraith, 1934:24). The sending of documents from one office to another is only

a small part of what is meant by “cooperation” and all the other things meant by the term would not necessarily go along with the transmission of documents. The indicators used are not sufficient for this purpose because we do not know, without looking at other evidence, how they might have been correlated with other indicators of cooperation. When Galbraith reconstructs the channels of communication, however, he is on sure ground.

The use of documents as indicators must not be confused with the use of facts or events as indicators. The events that took place on May 30, 1765, in the House of Burgesses, along with numerous other events throughout the colonies, are indicators of, let us say, the degree of tension between the colonies and England. But in order to know what took place on that day in the House of Burgesses, historians must first confront documents and then make inferences about events that are external to the documents. Only then can the events be used as indicators. And, of course, historians must go through similar steps in order to use events as correlates of other events. The use of external facts as correlates or as indicators, in other words, the making of inferences from established facts to other facts, is among the procedures that historians share with social scientists generally. In contrast, although the four types of inference from documents to events, which have been identified here, are sometimes found in the works of other social scientists, these four procedures are among the distinctive features of the historian’s craft.

In some cases, historians simultaneously use all of these four procedures in order to answer a single question. Testimony, social bookkeeping, correlates, and direct indicators are all used by Homans (1942) in order to establish the geographical location and boundaries between champion land and woodland in medieval England. Homans deals with four characteristics by which woodland and champion land differed from each other: size of fields, open or closed fields, certain agricultural practices, and the distribution of settlements. For the first of these characteristics, he relies on testimony. For the second he relies on testimony and on

a correlate, enclosure acts. For both the first and second, he also uses a direct indicator. If you go out and look you “can see that Devonshire, with its small squarish fields and big walls, has not the same landscape as Oxfordshire, though today the fields of both counties are, in the technical sense of the word, enclosed” (Homans, 1942:16). (This direct indicator would be even more telling if the book were written today, for the more sensitive eyes of aerial photography have since been used for this purpose.) For the third characteristic Homans cites documents of the social bookkeeping type, medieval surveys and extents. For the fourth, he again relies on testimony. While some of the evidence, when taken singly, may be called into question, the consistency with which these four types of evidence all point to the same conclusion answers all objections.

Although it is possible to classify neatly the evidence and procedures used by Homans in terms of the four categories presented here, it is not possible to do so in all cases, which the reader might think of. For example, those portions of the letters sent to Versailles by the French colonial governors in Quebec that purport to record information about events are a cross between testimony and social bookkeeping. One must read them in both lights. Perhaps such documents are common to similar situations, those in which one person reports to another in an official capacity but, instead of being one link in a complex and formally organized system of communication, is free to decide what he reports and how he reports it. Similarly, the distinction between correlates and indicators may not always be neat, and readers with experience in such matters can undoubtedly remember using procedures that fit into neither category. This is all to the good. For the methodological analysis of research situations which cannot be unambiguously classified in terms of the categories set forth here will not only define the limits within which these categories are useful, but will also force us towards greater clarity and precision. The further refinement of the categories presented here, and the identification of still further types of inference from documents to events, will provide historians

and all social scientists who use documents with a greater measure of self-conscious control over their materials and techniques.

The greatest historians may not need self-conscious control. Haskins did not need the general category “correlates.” Kosminsky did not need the category “social bookkeeping.” But few are blessed with their historical intuition. And among those who are not, only some can apply general methodological categories to the details of their specific problems. That is something, which the categories themselves cannot do for us. In short, while many may need methodological investigations such as this, perhaps only a few can make use of them. But if historians are willing to grant that there is any problem of methodology at all, then they must also welcome all steps toward methodological refinement.

#### NOTES

1. The phrase is from Gray, W. (1959). *Historian's handbook: A key to the study and writing of history* (p. 36). Boston: Houghton Mifflin.

2. The consistent application of this rule to early Spanish accounts of the Incas has led one author to a reconstruction of Inca society very different from those presented by earlier writers who had accepted the general descriptions given in the Spanish accounts. Cf. Moore, S. F. (1958). *Power and property in Inca Peru*. New York: Columbia University Press.

3. Decisions about authenticity are, of course, an example of “external criticism” and not of inferences from documents to events. Once the decision about authenticity is made, however, the fact of

authenticity or lack of authenticity is a characteristic of the document, which, in this case, is used as a correlate of external events.

4. This discussion has profited from Lazarsfeld (1959). So far as the writer knows, most historians become painfully aware of this fact—that the specific indicators of general concepts are not perfectly correlated with one another—only when they attempt to periodize. By the age of *laissez-faire*, or some such concept, historians refer to a thousand and one specific items. When they attempt to set the temporal boundaries of an age, however, they confront the fact that the specific indicators of the concept do not all change in equal degree with equal speed.

#### REFERENCES

- Chrimes, S. B. (1952). *An introduction to the administrative history of mediaeval England*. Oxford: Blackwell.
- Galbraith, V. H. (1934). *An introduction to the use of the public records*. Oxford: Clarendon Press.
- Haskins, C. H. (1918). *Norman institutions*. Cambridge, MA: Harvard University Press.
- Heckscher, E. F. (1955). *Mercantilism* (M. Shapiro, Trans., & E. F. Soderlund, Ed., Vol. 1, rev. ed.). New York: Macmillan.
- Homans, G. C. (1942). *English villagers of the thirteenth century*. Cambridge, MA: Harvard University Press.
- Kosminsky, E. A. (1956). *Studies in the agrarian history of England in the thirteenth century* (R. Kisch, Trans., & R. H. Hilton, Ed.). New York: Kelly & Millman.
- Lazarsfeld, P. F. (1959). Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3, pp. 476–543). New York: McGraw-Hill.



# 3.3

## POLITBURO IMAGES OF STALIN

NATHAN LEITES, ELSA BERNAUT, AND RAYMOND L. GARTHOFF\*

Hypotheses regarding differences (or lack of differences) in policy-orientation or in degrees of influence between the various members of the Soviet Politburo have always been of great interest to students of politics. Thus there have been frequent speculations regarding alleged differences in foreign policy lines and on the problem of succession. The absence of confirming or disconfirming data for any of these hypotheses is striking, and obvious in view of the secrecy that enshrouds the internal operations of the Politburo. Published statements of any kind by members of the Politburo have become infrequent in recent years. Such statements as are available for analysis have usually dealt with different subjects and have been made at different dates, so that they were difficult to compare from the point of view of testing hypotheses regarding differences in policy of influence.

Through Stalin's seventieth birthday, December 21, 1949, however, a rare opportunity for comparative analysis did occur. *Pravda* published articles by Politburo members Malenkov, Molotov, Beria, Voroshilov,

Mikoyan, Kaganovich, Bulganin, Andreyev, Khrushchev, Kosygin, and Shvernik (in this order), preceded by a joint message to Stalin from the Central Committee of the Party and the Council of Ministers of the USSR. These articles were reprinted in *Bolshevik*, the Party organ, and the Soviet press in general.<sup>1</sup> In addition, the anniversary issue of *Pravda* (but not *Bolshevik*) contained two articles on Stalin by persons who are not members of the Politburo, M. Shkiryatov (a Party Secretary) and A. Poskrebyshv (presumably Stalin's personal secretary), thus treating their statements on a par with those made by the members of the Politburo. This body of materials will be examined as to what it may reveal regarding the distribution of influence and attitudes within the Politburo.

While all the statements mentioned appear at first glance to express the same adulation of Stalin, they do contain nuances in style and emphasis. These nuances could more easily be dismissed as matters of individual rhetoric, of little relevance to political analysis, if the statements had been made by non-Soviet writers. But nuances in the political language

\*From Leites, N., Bernaut, E., & Garthoff, R. L. (1951). Politburo images of Stalin. *World Politics* 3:317-339.

used by members of the Politburo when talking about Stalin are of a different nature. Stalinism is not afraid of monotony and does not shun repetitiveness. Lack of complete uniformity of language is therefore possibly of political interest. It is worthwhile to examine the materials intensively in order to determine whether or not the differences in language, however subtle, fall into any patterns, and to explore the meaning of differentiations between groups or individuals in the Politburo. It seemed especially useful to approach the material with a view to investigating the degree of maintenance (or disuse and replacement) of earlier Bolshevik terms and themes.

Two major types of statements about the image of Stalin which can be discerned in the articles are analyzed in this paper. Table 1 gives the total frequencies of statements<sup>2</sup> concerning these ideas: first, Stalin in comparison to Lenin; and second, characterizations of

Stalin's dominant role, as "perfect Bolshevik" or "ideal Father." A third image, "Stalin" as person or symbol, is not presented in this table or discussed in detail because the difference between images is a more qualitative classification derived from analysis of the context within the articles; it is briefly discussed at the close of this article.

The frequencies of statements, when read across, indicate the weight given to "popular image" of Stalin. The articles were not uniform in length: Malenkov's article was approximately 3,500 words; those of Shvernik, Andreyev, Kosygin, Khrushchev, and Shkiryatov were each about 2,500 words; the others were each approximately 5,000 words. However, since the relative weight given to characterizations within each article is the subject of our attention here, no "weighing" of frequencies has been made in the table, and absolute figures have been used. . . .

**Table 1** References to Stalin in the Birthday Speeches December 21, 1949

<i>Politburo Member</i>	<i>Stalin: Lenin's Pupil or Equal?</i>			<i>Stalin: Perfect Bolshevik or Ideal Father?</i>		
	<i>Bolshevik Image</i>	<i>Ambiguous</i>	<i>Popular Image</i>	<i>Bolshevik Image</i>	<i>Ambiguous</i>	<i>Popular Image</i>
Molotov	5	1	0	12	3	0
Malenkov	4	0	2	11	3	0
Beria	13	3	1	15	1	2
Shvernik	4	0	5	2	2	2
Voroshilov	0	0	1	1	2	4
Mikoyan	2	2	9	3	0	5
Andreyev	1	0	2	3	0	15
Bulganin	0	1	6	0	0	3
Kosygin	1	0	3	0	0	8
Khrushchëv	0	1	4	0	0	7
Kaganovich	0	3	6	0	3	21
Shkiryatov	0	3	6	0	0	6
Poskrebyshev	2	1	3	0	0	10

### STALIN: LENIN'S PUPIL OR LENIN'S EQUAL?

In current Soviet public discourse, the "great" Lenin is not called "greater" than the "great" Stalin; nor is it affirmed explicitly that Lenin and Stalin are equal in "greatness." It is, however, possible to adopt formulations that suggest the former or the latter of these emphases.

In the articles on Stalin's birthday, the differences of stress fall into the pattern of tendencies toward what we have termed the "popular" and the "Bolshevik" images of Stalin; the popular image emphasizes Stalin's equality (and in some instances even primacy) in relation to Lenin, while the Bolshevik image lays more stress on Stalin as Lenin's "pupil," or the "continuer" of his work and ideas.

In the treatment of this point, the Bolshevik image characterizes the articles of the top group of the Politburo: Malenkov, Molotov, and Beria. In a "middle" position, using both images, are the joint article of the Central Committee and the Council of Ministers and Shvernik's article. Tendencies toward the popular image are expressed by Kosygin and Voroshilov (each of whom makes only two comparisons), Andreyev, and Poskrebyshev. The popular image is most frequently and clearly presented by Mikoyan, Kaganovich, Bulganin, Khrushchev, and Shkiryatov.

Beria uses the Bolshevik image, illustrated in the following examples, most frequently:

From his first steps of revolutionary activity Comrade Stalin stood unwaveringly *under* Lenin's banner. He was Lenin's true and devoted *follower*. He made his extremely valuable contributions to *Leninist* development of the Marxist Party's . . . tenets. . . . Establishing and developing *Leninism* and relying on *Lenin's instructions (ukazniya)*, Comrade Stalin developed the *tenets* of . . . industrialization. [*Digest:12*]

There are other instances where Beria states that "Comrade Stalin developed *Lenin's instructions*" (*Digest:13*) and "developed *Lenin's teaching* on the Party" (*Digest:12*), but this quotation is especially significant since Stalin is in Soviet writing almost

universally credited with the decision to collectivize and industrialize the country at a rapid tempo. There are many other references to Stalin's "arming the Party with Leninism," or "defending" or "advancing" Leninism, but these are not real comparisons.

There is one statement of equality on a situation (the conduct of the Civil War) concerning which Stalin has credited himself with a role possibly higher than Lenin's, so that equality in this respect would belong to the Bolshevik image.

During the difficult Civil War years *Lenin and Stalin* led the Party, the State, the Red Army and the country's entire defense. [*Digest:12*]

Beria even makes one statement about "the introduction of the *Leninist-Stalinist national policy*" (*Digest:13*) dealing with the one matter attributed to Stalin's own authorship prior to the middle twenties. Beria also mentions Stalin's investiture by Lenin, a theme that is rarely touched upon:

Lenin proposed that the Central Committee of the Party elect Comrade Stalin General Secretary of the Central Committee. Comrade Stalin has been working in this high post since April 3, 1922. [*Digest:12*]

As Lenin proposed, in 1923, that the Party consider the "removal" of Stalin from this "high post," Beria's reference is unusual. . . .

Molotov also expresses the Bolshevik image of Stalin in comparison to Lenin, emphasizing his theoretical continuation rather than personal discipleship, as Beria does. Both mention the fact that after Lenin's death, Stalin headed the Communist Party. Molotov goes on, however, to state:

Comrade Stalin upheld and developed *Lenin's theory* of the possibility of victory of socialism in one country. . . . [*Digest:7*]

As the . . . representative of creative Marxism, Comrade Stalin has highly developed the *Leninist* principles of strategy and tactics of our party. . . . [*Digest:10*]<sup>3</sup>

Molotov also expresses the Bolshevik image of Stalin as the successor to Lenin in his capacity as “head of the Party” and the preserver of its monolithic character, and says:

As the great *continuer* of the *cause* of immortal Lenin, Comrade Stalin stands *at the head* of all our socialist construction. . . . [Bolshevik:22]

Malenkov also stresses the Bolshevik image (despite two statements of apparent equality concerning their role in the Revolution):

Better than anyone else, Comrade Stalin profoundly *understood Lenin's* inspired *ideas* on a new-type Marxist party. [Digest:3]

A middle position, using both images frequently, is noticeable in the joint C.C.-Council of Ministers message, and in the article by Shvernik, entitled “Comrade Stalin—*Continuer* of the Great *Cause* of Lenin.” In addition to the title of his article, Shvernik makes three weaker Bolshevik image references to Lenin and Stalin, such as the one cited below.

From the first steps of his revolutionary struggle, Comrade Stalin was pervaded with a boundless faith in *Leninist* genius, and went on Lenin's path as the most loyal of his *pupils* and companions-in-arms. [Bolshevik:1]

On the other hand, he expresses the popular image four times, writing “together with Lenin, Comrade Stalin” (Bolshevik:91, twice), and “Lenin and Stalin” led the working class to victory (Bolshevik:91), and finally, in words borrowed from Mikoyan, he says: “Stalin—that is Lenin today” (Bolshevik:95).

Poskrebyshev (Stalin's secretary, and possibly a future member of the Politburo) also expresses a mixed attitude on this question, with three unequivocal statements of equality, three as “continuer of the cause of Lenin,” and two as teacher-pupil.

Kosygin, Andreyev, and Voroshilov employ the popular image more frequently than the Bolshevik but do not compare Lenin and Stalin often. Thus Kosygin writes: “The

ideas of *Lenin-Stalin* have triumphed. One-third of the population of the globe has entered firmly onto the path indicated by Lenin Stalin . . .” (Bolshevik:89), and later “path of socialism, indicated by *Lenin-Stalin* . . .” (Bolshevik:90). Kosygin even omits the name of Lenin in a passage where one might have expected to find it:

With the name of Stalin is indissolubly connected the creation of our Communist Party and of the first Soviet socialist state in the world. . . . [Bolshevik:86]

Just as Andreyev's article was predominantly devoted to agricultural matters, Voroshilov's article was concerned with military affairs, more specifically the strategy and conduct of the Great Fatherland War. In addition to two references to “the Party of Lenin and Stalin” he makes only one comparison, expressing equality.

During the years of the heroic struggle and labor [the Revolution], the Soviet people under the leadership of the Party of Bolsheviks, under the guidance of the great leaders *Lenin and Stalin*, secured a world-historical victory. [Bolshevik:35]

The popular image is clearly dominant, and frequent, in the articles of Mikoyan, Kaganovich, Bulganin, Khrushchev, and Shkiriyatov.

Thus, Mikoyan states:

*Stalin* not only fully mastered the entire scientific heritage of Marx, Engels and Lenin . . . [He “defended” and “brilliantly interpreted” it]; he also enriched Marxism-Leninism with a number of great discoveries, and further developed the Marxist-Leninist theory. In the words of Comrade Stalin Leninism is raised to a new, *higher* historical plane. . . . The Marxist-Leninist philosophy, which is transforming the world, has reached its *apex* in the works of Comrade Stalin. [Digest:19]<sup>4</sup>

Kaganovich is even more devoted to the use of the popular image, representing Stalin as equal to (or in rare instances even superior to) Lenin. There are no clear uses of the

Bolshevik image in his article, which abounds in comparisons.

Comrade Stalin did not *simply* defend and safeguard the Leninist theory of the possibility of the victory of socialism in one country, but on the foundation of rich experience of the struggle, he *creatively augmented* and *enriched* the theory. . . . [*Bolshevik*:59]

In one place Bulganin credits Stalin with the distinction between just and unjust wars (“as Stalin teaches . . .”), without any mention of Lenin, who first made this distinction, and until now has been generally so credited in the Soviet Union (*Bolshevik*:70).

Khrushchev also uses the popular image, with one possible exception, in all his comparisons of Lenin and Stalin. In addition to five references to “the X of Lenin and Stalin” (X = Party, teaching, idea, cause, and banner), he makes three statements of clear equality and one which may even attribute superiority to Stalin.

Herein lies Comrade Stalin’s tremendous and invaluable service. He is the true friend and *comrade-in-arms* of the great Lenin. [*Digest*:30]

. . . Stalin, who *together with Lenin* created the great Bolshevik Party, our socialist state, *enriched* Marxist-Leninist theory, and *raised it* to a new, *higher* level. [*Bolshevik*:80]

Shkiriyatov also expresses the extreme image most frequently, stating only three times that Stalin is continuing the “cause” or “banner” of Lenin, while using the phrase “the teaching of Lenin and Stalin” four times, and making six comparisons of Lenin and Stalin, in all of which they are clearly represented as equal.

Reviewing the treatment of this theme we see that there emerge rather distinctly a Bolshevik image and a popular image, in the treatment of the relative standing of Lenin and Stalin by Politburo members.

The “Bolshevik image” is most prominent in the articles of the top sector—Beria, Molotov, and Malenkov (in that order). It represents Stalin as the pupil of Lenin, his follower, and his continuer as Lenin’s successor,

who continued to implement, defend, and elaborate Leninism. He appears as the most loyal of Lenin’s followers and the one who best understood his ideas. Stalin is not considered as Lenin’s peer (with the single exception of Malenkov’s treatment of the October Revolution).

The “popular image” of Stalin is predominant, in varying degree, in the words of all the others, especially Kaganovich, Khrushchëv, Mikoyan, Bulganin, and Shkiriyatov. It represents Stalin as the equal of Lenin, also in situations where this was obviously not the case. In rare instances, Stalin even appears greater than Lenin.

---

#### STALIN: THE PERFECT BOLSHEVIK PARTY LEADER, OR THE IDEAL FATHER

---

The Bolshevik image is employed by Beria, Malenkov, Molotov, and to a lesser degree by Shvernik and Mikoyan. Stalin appears as the great “leader” and “teacher,” but by implication the Party is superior to him. He possesses a very high degree of Bolshevik virtues.

The perfect Bolshevik takes it for granted that his life is dedicated to the advancement of Communism, at whatever deprivations to himself. He regards it as improper to talk about ultimate values and personal sacrifices; attention, he feels, should be concentrated on discerning the correct line and carrying it through. The traits ascribed to Stalin by Beria, for instance, are almost all means to this end and are presented as such. A positive evaluation of a Bolshevik commends him for having made himself an effective tool in correct directions.

The popular image of Stalin, given much more profusely, does not present him as a Party leader impersonally fulfilling the moral obligation to render service to the proletariat by providing a correct policy line. It shows him as a People’s Leader in the Soviet Union and in the rest of the world, bestowing boundless paternal solicitude (*zabota*) on the “simple people.” The people, overwhelmed by surprise at finding such freely tendered goodness in one of their very own (*rodnoi*) on high, work harder and better for him in loving

gratitude. While the aim of the Party leader is to realize Communism in the future, at the cost of current hardships, the solicitude of the Leader of the People aims at satisfying human needs now. This he does, not only by laying down over-all policy, but also by innumerable concrete actions. In all this, Stalin possesses the virtues of an ideal father (sometimes brother and friend) which his children do not strive to equal. Stalin tends to become the creator of all good things.

The use of the Bolshevik image by the top group in this respect is far from excluding the use of elements of the popular image. Nevertheless, there is a differentiation, which we shall endeavor to show.

1. One of the aspects of the Bolshevik image of Stalin is his endowment with a very high degree of Bolshevik virtues. The implication is that these distinctive virtues should be emulated by less perfect Bolsheviks and that, although the chances of attaining Stalin's degree of perfection are slight, the model is clear, and there is no predetermined limit to advance.

For example, Beria says:

In Comrade Stalin the Soviet people saw even more clearly and distinctly the features of his great teacher, Lenin. They saw that our army and people were led into battle against a brutalized enemy by a tested leader who, like Lenin, was fearless in battle and merciless toward the enemies of the people; like Lenin, free of any semblance of panic: like Lenin, wise and bold in deciding complicated questions; like Lenin, clear and definite, just and honorable, loving his people as Lenin loved them. [Digest:15]<sup>5</sup>

Molotov also stresses Stalin's Bolshevik traits in several passages, (one) outstanding example follows below.

The works of Stalin are now appearing, containing his containing his works from 1901. It is impossible to overestimate the theoretical and political significance of this publication. Before our eyes, stage by stage, there unfolds the picture of the inspired creative work of the great Stalin, in all its diversity and spiritual wealth. Here, all the diverse practical questions of the work of the Bolshevik party and the international

communist movement and, together with this, complex scientific problems of history and philosophy are treated in the light of the ideas of Marxism-Leninism. . . . [Digest:9]

In most cases, popular image characterizations are admixed with Bolshevik statements showing Stalin as "leader" and "teacher." Of all the statements by the top group in the popular vein, only one (by Molotov) communicates a feeling or judgment by the speaker himself; all the other instances allege judgments or feelings of the *people*.

Comrade Stalin is rightfully considered a great and loyal friend of the freedom-loving peoples of the countries of people's democracy. . . . [Digest:3]

In addition to stressing his Bolshevik virtues, the Bolshevik image presents Stalin as leader in three forms: political strategist, teacher, and Party executive. We shall examine these in turn.

2. According to the Bolshevik image, Stalin's main role is to make a diagnosis and prognosis of the political situation and to derive the correct line from it. In the popular image of Stalin this is stressed much less. This aspect of the Bolshevik image is conveyed particularly by Molotov, as the examples below indicate.

. . . [Stalin's] ability . . . to show the Party the true way and to lead it to victory. [Digest:11]

In order that the anti-Hitler three-power coalition might be created during the war, it was necessary first to thwart the anti-Soviet plans of the governments of Britain, France . . . Comrade Stalin discerned in time the . . . Anglo-French intrigues . . . enabling us . . . to bring the developments of events to a point at which the governments of Britain and the U.S.A. were faced with the necessity of establishing an Anglo-Soviet-American . . . coalition. . . . [Digest:8]

3. Related to this in the Bolshevik image is Stalin's function in "teaching" the Party rules of organization, strategy, and tactics. This is another point less stressed in the

popular image of Stalin. But it is one of the main emphases of Malenkov (who may expect to take over this function). The following citations from his speech are but a few of many.

Comrade Stalin *teaches* that the Bolshevik Party is strong because . . . it multiplies its ties with the broad masses of the workers . . . Comrade Stalin *teaches* that without self-criticism we cannot advance . . . Comrade Stalin *teaches* that . . . Comrade Stalin *educates* the cadres of our Party. . . . [Digest:4–5]

Molotov and Beria emphasize Stalin's character as the "continuer," "defender," and "developer" of Leninism more than this teaching role, but they often do refer to Stalin as "leader and teacher." (This standard phrase is also found in the popular image but less frequently and prominently.)

4. The top-level statements frequently present as the major acting force not Stalin but the Party (or, sometimes) the "Soviet Union," or the "Soviet people," while other members of the Politburo stress the personal role of Stalin by omitting references to the Party. The Party is even credited with those services most often credited to Stalin by most of the others—inspiring, mobilizing, organizing. The term "leadership of the Party" clearly refers to others besides Stalin—in fact, to the speaker himself. Thus in the following passage, Malenkov mentions the Party eight times, and Stalin only once.

The friendship among peoples which is firmly established in our country is a great *achievement of the leadership* of the Bolshevik Party. Only the Bolshevik Party could forge the indissoluble fraternity among the peoples—the Bolshevik party which consistently carries forward the ideas of internationalism. . . . [The recent war] was a most serious one for the Bolshevik Party itself. The Party emerged from this test a great victor . . . following the instructions of Comrade Stalin, our Party constantly inspired the people and mobilized their efforts in the struggle against the enemy. The Party's organizational work united and directed. . . . Again the unsurpassed ability of the Bolshevik Party to mobilize the masses under the most difficult conditions was demonstrated. [Digest:4]

On the other hand, the image of Stalin as the People's Leader (the popular image) shows him acting directly, without using the transmission belt of the Party. Occasionally the "top group" and the "middle group" members use this image in topics intended for mass consumption:

. . . Stalin's voice in defense of peace . . . has penetrated throughout the world. . . . *All simple and honest people* responding to his appeal group themselves into powerful columns of fighters for peace. [Voroshilov, *Digest*:19; *Bolshevik*:44]

The popular image of Stalin, as we have indicated previously, does not stop at the limits which mark *the* Bolshevik characterization described above. Indeed, *it* very rarely uses them at all, except for casual and occasional reference to the standard term "leader" and "teacher." The articles of Kaganovich, Khrushchev, Shkiryatov, Poskrebyshev, Bulganin, Kosygin, and Andreyev, in roughly that descending order, are most expressive of the popular image, in the aspects presently under review. Mikoyan, and to a lesser degree Shvernik, also use it, but there are a number of mixed and even Bolshevik statements in their articles. On the other hand, the seven writers listed above have only four Bolshevik image statements in all their articles. Voroshilov is a special case; in his introduction and conclusion he makes a number of statements in the popular image.

1. In the popular image Stalin is characterized as the "father" of his people, who constantly helps them because of his "paternal solicitude" for them. (This is sometimes weakened to a "friend" relationship, and sometimes intimate relationship terms are not employed.) "The simple people" are grateful, loving, and industrious in return. For them Stalin is *rodnoi*, meaning "one's very own." and connoting familial intimacy.

Each of the members of the "bottom group" uses this description (to varying degrees, of course, as shall become evident). The following examples are by no means exhaustive of the instances used.

Kaganovich depicts Stalin in this manner in the following passages:

Comrade Stalin displays *exceptional solicitude* regarding miners and the alleviation of their labor. . . . The glorious army of railway workers responds to Comrade Stalin with *warm love, devotion*, and with a growing and improving transport [system] for his *paternal warmth and solicitude*. . . . The systematic increase of wages [etc.] . . . all these are the results of the *constant solicitude and attention of our very own [rodnoi]* Comrade Stalin, whom the people *lovingly* call *father and friend*. [*Bolshevik*:60–61]

Bulganin develops a similar image:

Comrade Stalin *always* displayed and displays up to the present time a *constant paternal solicitude* for the *bringing up* [*vyrashchivani*; used in the phrase “bringing up one’s children”] of military cadres, *educating* them in the spirit of supreme fidelity to the Bolshevik Party, in the spirit of self-sacrifice in the service of the people. . . . [*Bolshevik*:67]

Khrushchev similarly states:

Lenin and Stalin stood at the *cradle* of each Soviet republic, they guarded it from menacing dangers, *paternally* [*po-otcheski*] helped it to grow and become strong. . . . This is why all the peoples of our land, with the uncommon *warmth and feeling of filial love*, call the great Stalin *their very own [radnoi] father*. . . . [*Bolshevik*:81]

Andreyev, while not stressing this aspect of the extreme image, states:

*Attentively, paternally*, daily leading and *watching over* affairs on the collective farms . . . [is] Comrade Stalin. [*Digest*:29]

The two non-Politburo members, Shkiryatov and Poskrebyshev, both use this aspect of the popular image frequently. Poskrebyshev even titled his article “*Beloved Father and Great Teacher*.”

Shkiryatov writes:

The peoples of our country grow and become stronger like one *family*, and glorify Comrade Stalin—*father and friend* of all peoples of the USSR. [*Pravda*:11]

Stalin, our *father and friend*, instills in us a love for all that is *ours, native*—in science, in culture, in production, and *educates* into the Soviet people a *warm devotion* to its Motherland. . . . [*Pravda*:11]

2. As has already become evident, the popular image pictures Stalin as the People’s Leader, as contrasted to the emphasis on the Party and Stalin as Party leader in the moderate view. There are several aspects to being “People’s Leader,” and one which has been suggested in several of the quotations already cited shows Stalin as an opponent of “bureaucracy.” In his concern for the welfare of the simple people, he must overcome the indecency, selfishness, and malice of the bureaus standing between him and the people. Bulganin makes this almost explicit:

Comrade Stalin always paid great attention to the welfare of soldiers and sailors. He was interested in food standards, the quality of uniforms, and the weight of arms carried by soldiers. Comrade Stalin frequently pointed out in his orders that concern for the soldiers’ . . . welfare was the sacred duty of the commanders, that they must see to it most strictly that soldiers received all the food due under established standards, that the troops were given well-prepared warm meals in good time. . . . Due to the constant solicitude of Comrade Stalin for the suppliers of the troops our front fighters were well fed and comfortably and warmly clad. [*Digest*:28; *Bolshevik*:71]

Many other examples could be cited to demonstrate this aspect of the popular image.

The popular image of Stalin shows him, by implication, almost as a one-man Party-government-and-army apparatus. The previous quotations have pointed out this characterization of Stalin in situations where the welfare of the people required it. But this does not exhaust the range of his actions, and Kaganovich and Bulganin in particular extend Stalin’s active personal role to rather extreme lengths. According to Kaganovich:

. . . while . . . the countries of Europe, and the U.S.A., first of all, are slipping toward a crisis, here in the Soviet Union the socialist economy improves constantly. . . . We are obliged for this to the superiority of the socialist system of



economy, and *above all to Comrade Stalin's great energy, initiative and organizing genius.* [Digest:25]

Bulganin concerns himself with Stalin's role during the war, where Stalin performed an apparently prodigious amount of diverse labors constantly. Already in the Civil War,

Comrade Stalin was the creator of the most important . . . strategic: plans and the *direct leader* of the decisive battle operations. . . . At Tsaritsyn and Perm, at Petrograd and against Denikin, in the West against the Poland of Pans, and in the south against Wrangel—everywhere his iron will and military genius secured [*obespechivali*] the victory of Soviet forces. [Bolshevik:66]

And in the recent war,

*All operations* of the Great Fatherland War were planned by Comrade Stalin and executed under his guidance. There was *not a single operation* in the working out of which he did not participate. Before finally approving a plan . . . Comrade Stalin subjected it to thorough analysis and discussion with his closest [an unusual statement] . . . Comrade Stalin personally directed the whole course of every operation. Each day and even several times a day he verified the fulfillment of his orders, gave advice, and corrected the decisions of those in command, if there was need of this. [Bolshevik:66]

This image of Stalin as omnipresent and competent in every matter—an image never presented by the Politburo top group—is developed to a still further extreme by Poskrebyshev:

Attentively supervising the work of the leading *Michurinists* [the new geneticists], headed by Comrade Lysenko, Comrade Stalin gave them *daily assistance by his advice and instructions.* . . . Comrade Stalin must also be noted as a scientific innovator in specialized branches of science. . . . Among the old specialists in agriculture it was considered firmly established that the cultivation of citrus crops could not be extended on a wide scale in the region of the USSR Black Sea coast. . . . [Digest:34]

## STALIN: PERSON OR SYMBOL?

In our material, “Stalin” often refers to more than the man, J. V. Stalin. The boundary between references to Stalin the person and, as might be said, Stalin the symbol is blurred, probably on purpose. The top group, however, is more careful than the other to distinguish between these two images, and to lay stress on Stalin the person.

One way of indicating that Stalin is being referred to as a symbol is by speaking of his “name,” or actually declaring his name to be a “symbol.” Thus Beria states in his introductory paragraph:

Since the great Lenin there has been no *name* in the world so dear to the hearts of millions of working people as the *name* of the great leader, Comrade Stalin. [Digest:11]

And Molotov tells us that for “the world movement for peace”

. . . the name of Stalin is its great banner. [Digest:9]

Malenkov also states this:

The *name* of Comrade Stalin has long since become a *banner* of peace in the minds of the peoples of all countries. [Digest:3]

And Bulganin writes:

The *name* of Stalin became for the Soviet troops the *symbol of the greatness of our nation* and its heroism. They went into battle with the slogan: “For Stalin, for the Motherland!” [Bolshevik:71]

Another way of differentiating between Stalin the person and Stalin the symbol is by making explicit the *personal* character of the reference. In the birthday articles, Molotov, Shvernik, and Bulganin use this mode of expression most frequently. Although many other references which do not specify that Stalin the person is meant probably do mean this, the method remains, when used by the top group, an indication of instances where Stalin's

personal role is held to be highly significant. Malenkov uses a different method of achieving a similar effect. Although he refers to Stalin an average number of times (average number, 59 Malenkov's total, 60), a disproportionately large number of the references are to the effect that "Stalin teaches that . . ." or "as Stalin said," etc. Consequently, he says relatively less about other accomplishments of Stalin.

A technique used to transform "Stalin" from the person into the symbol is to employ the adjectival form of the word, "Stalinist." The Bolshevik image usually reserves the term "Stalinist" to describe the achievements of Stalin's regime rather than his personal accomplishments. The popular image is, on the whole, lax about this differentiation, and apparently allows personal and impersonal meanings to be given to "Stalinist," as well as to "Stalin."

The proportion of uses by the top group (Molotov, Beria, and Malenkov) of "Stalinist" as meaning "Stalin's" personally is only two out of a total of twenty-seven, in contrast to the very frequent use of the term in this meaning by all the others (excepting only Voroshilov's account of Stalin's role in the recent war). Very often statements are made describing "the Stalinist, Soviet path," "Soviet, Stalinist military science," and the like, inferring clearly that the term in these instances indicates merely "under the present regime" or "in a Bolshevik manner."

The relatively impersonal meaning of the adjective "Stalinist" is particularly evident in such passages as the following. Molotov, affirming that the Soviet Union has gained in strength over the last quarter of a century, says:

This is a very great service of Comrade *Stalin* and of *Stalinist leadership*. [*Digest*:6]

Presumably, "Stalinist leadership" here refers to Party leaders other than Stalin, and becomes a synonym for "Party." This is shown when Kaganovich, in a rare formulation, says:

A decisive condition for the victory of socialism was the incessant struggle of Comrade *Stalin* and of the united collective *Stalinist leadership* . . . for the realization of the general line of the Party. [*Bolshevik*:63]

## CONCLUSIONS

Two main conclusions emerge from this study of the birthday articles:

1. Despite many individual differences among these articles and despite the variations within each of them, two major images of Stalin may be constructed, toward which each article is oriented to its particular degree. Briefly, these images are Stalin the Party Chief and Stalin the People's Leader. The Party Chief is a very great man; the People's Leader stands higher than any man. The Party Chief is characterized by Bolshevik traits; the People's Leader by constant and boundless solicitude for the welfare of all. We have referred for the sake of brevity to the first as "*the Bolshevik image*," and to the second as "*the popular image*."

2. Three groups within the Politburo can be distinguished in terms of using these images. Malenkov, Molotov, and Beria, who presumably are the most influential members of the Politburo, stress the Bolshevik image of Stalin more than the other members, although indications of the popular image are not totally absent from their statements. Kaganovich, Bulganin, Khrushchev, Kosygin, and to a lesser degree Mikoyan and Andreyev, occupy positions near the popular image (as do Shkiryatov and Poskrebyshv). Shvernik and the joint Party-government address occupy a middle position. Voroshilov is a special case, presenting the popular image of Stalin in his introduction and peroration, but a very moderate Bolshevik image in terms of specific military operations (in contrast to Bulganin).

These two images of Stalin can now be reviewed with two questions in mind: (1) To whom is either image addressed? Is there a preferred audience for the popular image and another such audience for the Bolshevik image? (2) What political significance can be attached to the finding that the Bolshevik image is stressed by the "top group" in the Politburo, while the popular image is used most freely by the "bottom group"?

Concerning the first question, it should be remembered that all statements analyzed in this paper were published; they were not made

in private. As public statements they were not primarily, or at any rate not exclusively, addressed to Stalin. It is reasonable to assume that the “masses” of the Soviet population were meant to be the consumers of the popular image, whereas the Bolshevik image was offered primarily for adoption by Communists, i.e., a small segment of the population. It is characteristic of Bolshevism, though paradoxical to Western thinking, that the symbols of nearness and intimacy (“father,” “solicitude,” etc.) appear most frequently in the popular image of Stalin and are stressed for that audience which is far removed from Stalin. Those closer to Stalin politically are permitted to speak of him in terms of lesser personal intimacy (“leader of the party”). This paradox results partly from the merely instrumental use in Bolshevik language of words indicating personal nearness, and partly from the Bolshevik deprecation of such nearness in political relationships. The ideal Party member does not stress any gratification he may derive from intimacy with others, much as he may use such intimacy for political ends.

For this reason it is difficult to answer the second question with certainty. It cannot be ruled out that the Politburo—or a leading group within it, or Stalin personally—decided to use both images of Stalin in the birthday statements and to adopt a certain distribution of roles among its members in presenting them. (Such a decision may have taken the form of an editorial scrutiny of each statement, in the course of which the differentiation of language was imposed.)

However, the assumption that there was a decision within the Politburo on the use of different images of Stalin does not preclude certain tentative conclusions about the status of the groups within the Politburo. The emphasis on the Bolshevik image by a few members of the Politburo and on the popular image by others not only reflects the Bolshevik evaluation of the Party as distinguished from, and superior to, the masses at large, but also indicates the relative distance of the speakers from Stalin. In the situation under review, it is a privilege for a member of the Politburo to refrain from using the crudest

form of adulation, words signifying personal intimacy and emotions; that is, private, rather than political, words. Given the Bolshevik evaluation of political as against private life, the use of the Bolshevik image indicates higher political status. Hence, a planned distribution of roles in using the two images of Stalin on the occasion of his birthday would still indicate a political stratification of the Politburo, though not necessarily political antagonism within it.

Unless one were to make the somewhat absurd assumption that the roles to be performed on this occasion were distributed by lot, or the improbable assumption that they were assigned for the purpose of concealing the real stratification within the Politburo, those members who stress the Bolshevik image could be assumed to be politically closer to Stalin than those who do not.

The assumption that there had been a decision of some kind on the use of the two images would appear more plausible if either image were used by certain members of the Politburo without the admixture of elements taken from the other. As it is, the difference between the “top group” and the “bottom group” is one of *emphasis* in imagery. For this reason, we are inclined to regard the differentiation of political language discussed in this article as the result of individual choices rather than of a central decision. However, in this case we may assume that the stress—whether conscious or not—of any given Politburo member on the one or the other image of Stalin was related to his status in the Politburo in the fashion indicated above.

## NOTES

1. As far as is feasible, quotations are given from the translation in Volume 1, No. 52, of *The Current Digest of the Soviet Press* (hereafter cited as *Digest*). Other passages have been translated from the December 1949, No. 24 *Bolshevik*. All italics, unless otherwise indicated, are by the authors of this article.

2. A “statement” for the purposes of this table, means each incidence of an explicit idea, and may vary from a phrase to a paragraph. The examples cited in the text should clarify this point.

3. December 21, 1929, Molotov wrote more specifically that Stalin had been a “man of practice” (*praktik i organizator*) up to Lenin’s death, after which he became a “theoretician.” Even in 1919 Molotov has not quite suppressed his tendency to deny that Stalin was manifestly perfect from the start. He begins his speech by saying: “It is *now* particularly clear how very fortunate it was . . . that after Lenin the Communist Party of the USSR was headed by Comrade Stalin” (*Digest*:6). In the Bolshevik atmosphere of veiled language, this is bound to be understood, to some extent, as conveying: It was not *always* clear.

4. Although Molotov and Beria both praise Stalin as the theorist, they do not state explicitly (or clearly implicitly) that Stalin is as great a theorist as Lenin, to say nothing of the statement that “Marxist-Leninist philosophy has reached its apex” in Stalin’s work.

5. “Loving the people” also belongs to the popular image. These occasional popular image terms in a moderate picture may be the effect of reverse seepage of esoteric propaganda into the constantly assaulted esoteric integrity of the top group.

# 3.4

## QUANTITATIVE AND QUALITATIVE APPROACHES TO CONTENT ANALYSIS

ALEXANDER GEORGE\*

Researchers have long debated the respective merits and uses of “quantitative” and “qualitative” approaches to content analysis. . . . Most writers on content analysis have made quantification a component of their definition of content analysis. In effect, therefore, they exclude the qualitative approach as being something other than content analysis.

Quantitative content analysis is, in the first instance, a statistical technique for obtaining descriptive data on content variables. Its value in this respect is that it offers the possibility of obtaining more precise, objective, and reliable observations about the frequency with which given content characteristics occur either singly or in conjunction with one another. In other words, the quantitative approach substitutes controlled observation and systematic counting for impressionistic

ways of observing frequencies of occurrence.<sup>1</sup> The term “qualitative,” on the other hand, has been used to refer to a number of different aspects of research . . . :

1. The preliminary reading of communications materials for purposes of hypothesis formation and the discovery of new relationships

*As against*

Systematic content analysis for purposes of testing hypotheses.

2. An impressionistic procedure for making observations about content characteristics

*As against*

A systematic procedure for obtaining precise, objective, and reliable data.

3. Dichotomous attributes (i.e., attributes, which can be predicated only as belonging or not belonging to an object)

\*Excerpt from George, A. L. (1959). Quantitative and qualitative approaches to content analysis. In I. de Sola Pool (Ed.), *Trends in content analysis* (pp. 7–32). Urbana: University of Illinois Press.

*As against*

Attributes which permit exact measurement (i.e., the true quantitative variable) or rank ordering (i.e., the serial).

4. A “flexible” procedure for making content-descriptive observations, or “coding” judgments

*As against*

A “rigid” procedure for doing the same.

## FREQUENCY AND NON-FREQUENCY CONTENT INDICATORS

---

While these four distinctions are important in themselves, they do not serve to differentiate between the two approaches to the analysis of communication. . . . We therefore introduce a somewhat different distinction, which focuses on the aspects of the communication content from which the analyst draws inferences regarding non-content variables.

1. Quantitative content analysis, as we here define it, is concerned with the *frequency* of occurrence of given content characteristics; that is, the investigator works with the frequency of occurrence of certain content characteristics.

2. Inferences from content to non-content variables, however, need not always be based on the frequency values of content features. The content term in an inferential hypothesis or statement of relationship may consist of the mere *presence* or *absence* of a given content characteristic or a content syndrome within a designated body of communication. It is the latter type of communication analysis, which makes use of “non-frequency” content indicators for purposes of inference, that is regarded here as the non-quantitative or non-statistical variant of content analysis.

The distinction we have introduced concerns the type of content indicator utilized for purposes of inference. Given (different uses of) . . . “quantitative” and “qualitative,” it is desirable to introduce a new set of terms. We employ the term “non-frequency” to describe

the type of non-quantitative, non-statistical content analysis, which uses the presence or absence of a certain content characteristic or syndrome as a content indicator in an inferential hypothesis. In contrast, a “frequency” content indicator is one in which the number of times one or more content characteristics occur is regarded as relevant for purposes of inference.

The distinction between frequency and non-frequency analysis, it should be noted, is independent of the aforementioned four dimensions to which the terms quantitative and qualitative are sometimes applied. Thus, both in frequency and non-frequency analysis (one can distinguish) between hypothesis-formation and hypothesis-testing phases of research, between impressionistic and systematic types of content description, between flexible and rigid procedures for making content-descriptive judgments.

Nor is the familiar distinction in the theory of measurement between dichotomous, serial, and quantitative *attributes* equivalent to the distinction advanced here. Thus, . . . frequency as well as non-frequency analysis may be concerned with dichotomous attributes, that is, attributes which can be predicated only as belonging or not belonging to an object. This is the case, for example, in the simple word-counts . . . deciding whether a certain word or symbol (“democracy,” “Germany,” “Stalin”) does or does not appear in each sentence, paragraph or article. . . .

[T]he difference between the two approaches is that frequency analysis, even when it deals with dichotomous attributes, always singles out frequency distributions as a basis for making inferences. In contrast, the non-frequency approach utilizes the mere occurrence or nonoccurrence of attributes . . . for purposes of inference. Thus, for example, [from] a quantitative study, which shows a sharp decline in number of references to Stalin in *Pravda*, the frequency analyst might infer that the successors to Stalin are attempting to downgrade the former dictator or are trying to dissociate themselves from him. On the other hand, the non-frequency analyst might make a similar inference from the fact that in a public speech one of Stalin’s

successors pointedly failed to mention him when discussing a particular subject (e.g., credit for the Soviet victory in World War II) where mention of Stalin would formerly have been obligatory. In one case, it is the frequency distribution of attention to “Stalin” over a period of time on which the inference rests. In the other, it is the mere occurrence or nonoccurrence of the word “Stalin” on a particular occasion,\* which serves as a basis for the inference. Yet, in both of these examples, the investigator deals with a dichotomous attribute . . . the presence or absence of “Stalin” in a given unit of communication.

Furthermore, the use of frequency and non-frequency methods is not determined by the fact of multiple or single occurrence of the content feature in question within the communication under examination. The fact that a content feature does occur more than once within a communication does not oblige the investigator to count its frequency. The important fact about that content feature for his inference may be merely that it occurs at all within a prescribed communication.

It should be noted, finally, that the non-frequency approach to content analysis is really an older and more conventional way of interpreting communication and drawing inferences from it than is the quantitative approach. The resemblance of the non-frequency approach to traditional methods of textual analysis, moreover, will become obvious when we consider some of the characteristics of this approach.

#### SOME EXAMPLES OF NON-FREQUENCY CONTENT ANALYSIS

[Two] examples will illustrate the nature of the non-frequency approach and some of the disadvantages of relying exclusively upon frequency or quantitative content analysis. The examples are drawn from wartime propaganda analyses of German communications made by personnel of the Analysis Division, Foreign Broadcast Intelligence

Service, and Federal Communications Commission (FCC).

1. [T]he FCC analyst inferred that the Nazi Propaganda Ministry was attempting to discourage the German public, albeit indirectly, from expecting a resurgence of the U-boats. This inference was also based in part upon a non-frequency content indicator. Hans Fritzsche, the leading radio commentator, had asserted the following in discussing a recent success achieved by German U-boats: “. . . we are not so naive as to indulge in speculation about the future on the basis of the fact of this victory . . .” In focusing upon this statement, the FCC analyst was not concerned with the frequency of the theme in Fritzsche’s talk or in other German propaganda accounts of the same U-boat “victory,” or with the question of whether it now appeared more or less frequently than in earlier propaganda on the U-boats. For his purpose, it sufficed that the content theme was present even once in the context of Fritzsche’s remarks about the latest German U-boat victory.

It is interesting to speculate on what would have happened had a frequency (quantitative) approach been employed in this case. In the first place, it is problematic whether a content category could have been set up that would have precisely caught the meaning of this one phrase in Fritzsche’s talk. Secondly, since the phrase (or its equivalent) appeared at best only a very few times in German propaganda at the time, the propaganda analyst, in looking over the quantitative results, might well have dismissed it as a “minor theme” or lumped it together with other items in a “miscellaneous” or “other” category. In other words, if this single phrase from Fritzsche’s talk had been subsumed under a frequency indicator, it might well have lost its inferential significance.

2. The (second) example is quite similar. [Towards the end of World War II,] Mussolini had set up a Republican-Fascist government following his “liberation” by German parachutists (from his imprisonment). German propaganda gave quite a play to these events

\*In the context of what is expected, a known norm.

and celebrated Mussolini's re-establishment of a pro-Axis Italian government. In looking over this propaganda, the FCC analyst noted that, after a few days, a minor theme of some sobriety was introduced into the otherwise enthusiastic publicity on Mussolini and his new government. Only a few Nazi papers carried the new message and, where it did appear, it was rather well hidden. For example, in the *Völkischer Beobachter*, September 9, 1943, the following sentence appeared: "... the battle is not yet won by the changes proclaimed by Mussolini, and the structural changes undertaken by him must not be regarded as a guarantee of future greatness."

The significance of this new theme to the FCC analyst lay in the fact that it appeared at all. In other words, he made use of it as a non-frequency content indicator. Had this new theme been subsumed under a general frequency-type indicator in a quantitative study, it would probably have passed unnoticed. But when singled out as a non-frequency indicator, the theme, although repeated only a few times in the total Nazi propaganda on Mussolini, provided the basis for an important inference. The FCC analyst inferred, from the appearance of the theme, that the Propaganda Ministry had decided to moderate the public's expectations regarding a resurgence of Italian fascism. Continuing the chain of inference, the FCC analyst then reasoned that such a propaganda goal must have been adopted as the result of a new, more sober estimate by Nazi leaders of the potential of Mussolini's new government. This inference was subsequently verified [by] material appearing in *The Goebbels Diaries* (Lochner, 1948).

#### SOME DIFFICULTIES IN APPLYING QUANTITATIVE CONTENT ANALYSIS FOR THE STUDY OF INSTRUMENTAL ASPECTS OF COMMUNICATION

[I]n such fields as clinical psychiatry and propaganda analysis, content analysis is often used as a diagnostic tool for making causal

interpretations about a single goal-oriented communication. In order to identify and explore some of the special problems that arise in this type of content analysis, we shall further examine the case of propaganda analysis. Other communication analyses, which operate within the framework of an instrumental model, may encounter similar problems. We will discuss the following: (1) the problem of coding irrelevant content, (2) the problem of changes in the speaker's strategy, (3) the problem of an expanding universe of relevant communication, and (4) the problem of structural characteristics of instrumental communication.\*

These problems arise in part from the characteristics of propaganda communication, and in part [from] the investigator's interest in making specific inferences about some aspect of the communicator's purposive behavior. In any case, the result is that a considerable portion of the research effort must be given to discovering new hypotheses or refining old ones; systematic quantitative analysis for purposes of testing inferential hypotheses is often difficult, infeasible, or unnecessary; and, finally, non-quantitative (non-frequency) content indicators are often more appropriate and productive than quantitative (frequency) indicators.

#### The Problem of Coding Irrelevant Content

A variety of specific goals and strategies are usually pursued in propaganda communications. The propaganda analyst, however, may be interested in making inferences only about one or a few matters of policy interest. Accordingly, he must exercise care in considering which passages in the stream of communication are *relevant* to each of the goals or strategies of the communicator.

The difficulty of arriving at such judgments of relevance and the considerable sensitivity and discrimination, which are required for this purpose, are often reasons for not undertaking elaborate quantitative "fishing

\*These issues are just as relevant today in light of the use of content analysis techniques on the Web.



expeditions” in any sizable body of propaganda communications. Not all the individual items tabulated under any given content category in such a “fishing expedition” may be relevant to the specific inference, which the analyst would like to make about the speaker’s state of mind.

We are referring here obliquely to one of the important requirements of statistical content analysis, namely, that it be “systematic” in the sense that “*all* of the relevant content . . . be analyzed in terms of *all* of the relevant categories, for the problem at hand” (Berelson, 1952:17).

But the obverse of this requirement—that none of the *irrelevant* content be analyzed—is equally important and is a weighty reason for not undertaking elaborate quantitative content analyses of the “fishing expedition” variety. . . . In some cases the inclusion of irrelevant content in the analysis may be no more than a waste of manpower. But in other cases, it may rule out the possibility of making a useful inference or lead to wholly mistaken inferences. The problem may become particularly acute when the investigator, engaged in a “fishing expedition” of this sort, deliberately selects broad content categories in order to ensure large enough frequencies for purposes of subsequent statistical analysis.\*

The danger of coding irrelevant content is minimized when research is designed to test clear-cut hypotheses. Hypotheses usually indicate or imply the realm of relevant content or the appropriate sample to be coded.

However, precise hypothesis formation—the assertion of a relationship between a content indicator and one or more communicator variables—is often difficult in propaganda analysis. This difficulty reflects the rudimentary state of the scientific study of communication. The lack of good hypotheses about relationships between content variables and communicator variables makes it difficult for the propaganda analyst to circumscribe the terms and categories for specific investigations.

This difficulty, of course, is by no means confined to propaganda analysis. . . . In a sober assessment of the results of their large-scale study of symbols as indices of political values, attitudes, and ideological dispositions, [Lasswell lamented]:

. . . there is as yet no good theory of symbolic communication by which to predict how given values, attitudes, or ideologies will be expressed in manifest symbols. The extant theories tend to deal with values, attitudes, and ideologies as the ultimate units, not with the symbolic atoms of which they are composed. There is almost no theory of language that predicts the specific words one will emit in the course of expressing the contents of his thoughts. Theories in philosophy or in the sociology of knowledge sometimes enable us to predict ideas that will be expressed by persons with certain other ideas or social characteristics. But little thought has been given to predicting the specific words in which these ideas will be cloaked. The content analyst, therefore, does not know what to expect. (Lasswell, Lerner, & Pool, 1952:49)

In summary, there are relatively few theories or general hypotheses about symbolic behavior available for testing by means of rigorous quantitative content analysis. . . . [S]ome investigators, [therefore] employ quantitative content analysis for purposes of a “fishing expedition”; large quantities of content data are collected without guidance of clear-cut hypotheses in the hope of discovering, at the end of the study, new relationships and new hypotheses. Such studies tend to be time consuming, wasteful, and generally unproductive. Disappointing results with “fishing expeditions” are particularly likely when large quantities of material are processed (aided by) clerical personnel to do the coding. As a result (of being locked into a fixed coding instrument), there is insufficient opportunity to refine categories and it is usually not possible to recode the bulky material as many times as necessary in order to produce content data appropriate for testing interesting hypotheses.

\*And thereby failing to record the distinctions that may prove critical in supporting the desired inferences.

### The Problem of Changes in the Speaker's Strategy

Due to the circumstances, which have been described, the "qualitative" phase of hypothesis formation may properly receive unusual emphasis in propaganda analysis. [This is justified by] the fact that the propagandist's strategy on any single subject may change abruptly at any time. In attempting inferences about the speaker's state of mind, . . . the analyst cannot easily draw up a set of content categories, which will be appropriate for all possible shifts in the communication strategy of the speaker. [He] will hesitate to commit himself to systematic quantitative description because he fears that the speaker's strategy may change while the count is being made. If such a change is unnoticed . . . , the value of the results of the quantitative tabulation . . . may be lost. For then, such content data might well be ambiguous or inappropriate for purposes of inference.

In propaganda analysis, the instrumental use to which communication is put by the speaker is regarded as a highly unstable variable, which intervenes between various other antecedent conditions of communication (e.g., speaker's attitude and state of mind, the conditions and calculations, which have affected choice of action) and the content variable itself. In this respect, propaganda analysis has much in common with the analysis of psychotherapy protocols. Both the propaganda analyst and the psychotherapist are sensitive to the possibility that the communication intention and strategy of the speaker can change frequently during . . . a systematic count of the content features of what he says. . . . [E]xcept when there is reason to

believe that the content features selected as indicators are insensitive to variations in the speaker's strategy, frequency counts may be inappropriate as a means of inferring the speaker's attitudes, state of mind, and . . . conditions that have influenced his choice of a communication strategy or goal.

In propaganda analysis, typically, the investigator is interested in inferring one or more of the following antecedent conditions of the propagandist's communication: his propaganda goals and techniques; the estimates, expectations, and policy intentions of the leadership group for whom the propagandist is speaking which have influenced the adoption of a particular propaganda strategy; the situational factors or changes which have influenced the leadership's estimates, expectations, and policy intentions and/or the propagandist's choice of communication goals and techniques.

Investigators interested in inferring elite estimates, expectations, policy intentions, and/or situational factors, which lay behind the adoption of a particular propaganda goal or strategy may employ one of two rather different methods of inference. They can attempt to find content indicators which directly reflect the component of elite behavior or the situational factor in which he is interested, or he can attempt first to infer the speaker's propaganda goal and then proceed step by step to account for the selection of that goal in terms of elite estimates, expectations, policy intentions and/or situational factors.

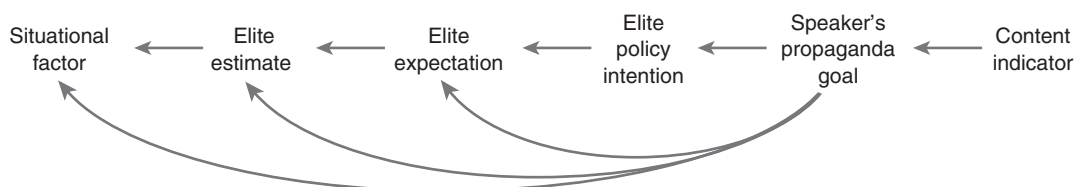
The first of these two methods of inference bypasses consideration of the speaker's propaganda strategy. The inferences made with this *direct method* are one-step inferences, as follows:



Figure 1

In contrast, the *indirect method* is comprised of an inferential chain of two or more steps, the first of which is always an inference

about the speaker's goal or strategy. It may be depicted, in somewhat simplified form, as follows:



**Figure 2**

The direct method . . . can be successfully employed only if content features can be found which occur *regularly* and *only* when a certain type of elite policy intention, expectation, estimate, or situational factor occurs. The types of regularities or generalizations that the direct method requires as a basis for inferences, therefore, are correlations of a non-causal character. It is important to recognize that the content terms in such correlations must be insensitive to possible variations in propaganda strategy. This is necessary because propaganda strategy is an intervening and relatively unstable variable between elite policy behavior and propaganda content. The direct method is on firm ground only when it employs as content indicators features of the communication over which the propagandist does not exercise control or of whose information-giving value regarding elite policy behavior he remains unaware. Such content features are likely to be symptomatic features of a propagandist's behavior rather than part of his communication intention.<sup>2</sup>

The indirect method, on the other hand, attempts to utilize for purposes of inference the fact that the behavior of the propagandist in selecting communications goals and strategies constitutes an intervening set of events between elite policy behavior and the dependent variable (content of propaganda). Therefore, the investigator who employs the indirect method attempts to identify content features in the propaganda, which are sensitive to and dependent upon the speaker's strategy.

The distinction between content features that are sensitive and insensitive to variations in the

speaker's strategy is useful not only in propaganda analysis, but whenever content analysis is used on instrumentally manipulated material. In such cases, obtaining good indices of attitude, value, etc., would seem to require either the avoidance of content features, which are likely to be sensitive in the first instance to variations in the speaker's communication strategy or a sophisticated awareness of that strategy.

### The Problem of an Expanding Universe of Relevant Communication

Another characteristic of propaganda, which has procedural implications, is the fact that the universe of relevant communication may be expanding while [its] analyst is attempting to draw inferences from it. . . . [Under these conditions], the propaganda analyst finds himself trying to keep up with the flow of communication that has some relevance to his problem. . . . [A]s new statements on the topic are made by the source, . . . the set of alternative hypotheses under consideration and [corresponding] content categories [have to be revised]. . . .

These circumstances frequently rule out quantitative content description. A familiar prerequisite of quantitative content analysis is that the investigator knows what he is looking for before beginning to count. The propaganda analyst (who relies on quantitative accounts of communications) . . . cannot be confident that the data provided will still be adequate for purposes of inference when new statements on the topic are received from the source. For the most

recent communication may throw new light on the inferential problem and, based on these new insights, the propaganda analyst may have to reread and reinterpret the earlier propaganda communications in their original form. . . .

A similar problem arises, we may note, when information about *non-content* events bearing upon the inferential problem comes to the attention of the propaganda analyst *after* he has received and analyzed the relevant propaganda communication. Such non-content events may permit the analyst retrospectively to formulate more discriminating hypotheses about the inferential significance of the propaganda. And, for this purpose, it may be necessary for him to reread and reappraise the propaganda communication in its original form, in the light of the new information available on relevant non-content events.

### The Problem of Structural Characteristics of Instrumental Communication

Propaganda analysis procedures are much influenced, finally, by the necessity to take into account the structure of individual propaganda communications. Different structural types of communication are encountered in the flow of propaganda available for analysis. An article by Goebbels appearing in *Das Reich*, for example, was structurally different from a speech by Hitler; and both, certainly, were structurally different from German radio news broadcasts.

The propaganda intention of an individual communication (and its effect as well) often depends not merely on the explicit content of the individual statements or propositions therein contained but also upon the structural interrelationship of these statements within that communication.

Thus, what may be called the “whole-part” problem in content analysis has several important procedural implications. It may affect [the] choice of counting units and categories as well as the decision on the type of content indicator (frequency or non-frequency) to be employed.

Awareness of the “whole-part” problem often leads the propaganda analyst to be critical of an important implicit assumption of statistical content analyses, namely, that each individual item counted as falling under a designated content category is of equal significance for purposes of inference.<sup>3</sup> Similarly, the propaganda analyst is often critical of the assumption that the inferential significance of explicit propositions, themes, or statements is dependent upon the precise frequency of their occurrence. Rather, he may find explicit propositions of significance for purposes of ascertaining the strategy of the propagandist because they occur at all or because they occur in a certain relationship to each other within the communication.

This does not mean that frequency counts are useless for purposes of propaganda analysis. Frequency tabulation of words, clichés, stereotypes, and slogans may provide an indication of propaganda emphasis and techniques as well as of intentions. But such tabulations in themselves give no clue to the meaning of the content in question. They are of value, therefore, only when the investigator has prior or independent knowledge of their meaning, role, and significance in the system of language habits under study.<sup>4</sup>

. . . The procedure employed in ascertaining the propositional content of a propaganda communication and in weighing the structural interrelationships of parts therein undoubtedly is often less systematic than in rigorous quantitative content analysis in which coding judgments are closely prescribed. But in principle, the reliability of such content observations, too, is subject to investigation.

### SOME CHARACTERISTICS AND SPECIAL PROBLEMS OF THE NON-FREQUENCY APPROACH

The preceding discussion has already suggested some of the characteristics of non-frequency analysis of instrumental communications. In this section, we recapitulate these characteristics briefly and single out for more extended comment the special problems to which they give rise.

### Selection of Content Categories: The Search for Specific Discriminating Categories

In some quantitative investigations, the technical requirement of relatively large numbers for statistical analysis appears to exercise an important influence on the choice of content categories and on the size of the sample of raw material to be coded. Symbols and themes with low frequency of occurrence may be either ignored or grouped together under broader content categories.

The conscious selection of content categories and sample size with an eye to satisfying technical requirements of statistical analysis may be justified when the research objective is to make general inferences. But such a criterion is inappropriate when, as in propaganda analysis, the object is to make specific inferences about events at particular times and places. In the latter case, valuable opportunities for making inferences are lost. . . .

The investigator who is aware of the value of non-frequency indicators tries, rather, to formulate ever more discriminating content categories. He deliberately attempts to “narrow down” the categories and to make them relatively more specific. The fact that this results in low frequencies, in a single occurrence, or in no occurrence at all of the content feature in question is not of concern to him since he expects to employ a non-frequency indicator for purposes of inference. . . .

### Emphasis on Hypothesis Formation as Against Hypothesis Testing

Perhaps more so than in most frequency analyses, the investigator who employs the non-frequency approach gives unusual attention and effort to the hypothesis-formation phase of research. There are a number of reasons for this emphasis to which we have already alluded: the search for more discriminating categories, the need to exclude irrelevant content, and, of course, the rudimentary state of knowledge and theory about the relationships between content and communicator variables.

### Relative Emphasis Upon Validity as Against Reliability of Semantical Content Description

The non-frequency approach places more emphasis upon obtaining valid estimates of the speaker’s intended meaning than do many versions of quantitative content analysis. Because he usually deals with relatively large frequencies, the quantitative investigator can work with somewhat lower validity requirements and can (and must) pay greater attention to reliability considerations. The inclusion of a small number of incorrect determinations of the speaker’s intended meaning under a content category composed of large frequencies will probably not affect the final analysis greatly. In contrast, just because he works with low frequencies or single occurrences, the non-frequency analyst cannot afford to risk making any invalid determinations of the speaker’s intended meaning.

Given the crucial importance in non-frequency analysis of validly estimating the one or few meanings, which may be of inferential significance, the investigator concentrates upon making an intensive assessment of contextual factors upon which such meanings are likely to depend. This type of assessment, however, is particularly difficult to objectify, for it requires taking into account the situational and behavioral, as well as the linguistic contexts of given words. Accordingly, the procedure for inferring the speaker’s precise intended meanings cannot easily be made fully explicit. Investigators who attempt to infer intended meanings must usually settle for relatively flexible and interpretative procedures of coding content.

Each judgment of an intended meaning is a separate inference arrived at by taking into account not only dictionary meanings and rules of the language, but also all relevant aspects of the context: situational and behavioral as well as linguistic. The concern with inferring intended meaning in this manner does not distinguish the non-frequency approach from all frequency analysis. But it does serve to differentiate it radically from

one variant of the quantitative approach known as “manifest” content analysis.\*

... [I]n “manifest” content analysis the investigator estimates the meanings of words by applying a set of external criteria as to the usual, customary, or most frequent meaning of the words in question.<sup>5</sup> Such a judgment or estimate of meaning is not [situation] specific. . . . [Use of such criteria] increases the objectivity of the content-descriptive procedure and facilitates achieving reliability of results, but . . . may seriously prejudice the validity of results if the intended meanings . . . differ . . . from the meanings which those words ordinarily have.<sup>6</sup>

In coding content for its usual, or “manifest” meaning, the investigator needs to be familiar with the general rules of the language, the customary meanings of words for all users of that language, and—in some varieties of quantitative content analysis—with the usual or most frequent language habits of the communicator.

... [T]o make valid inference of intended meaning in each specific instance of communication, the investigator also takes into account the situational and behavioral contexts of that communication. He does so in order to determine which of the possible meanings of the words in question the speaker intends to convey in the instance at hand and the precise shading of his intended meaning. . . .

In taking into account the *behavioral* context of words, the investigator considers the instrumental aspect of the communication in its broad action setting. In order to interpret the precise meaning intended by the speaker in any individual instance he takes into account the *purpose or objective, which the specific communication is, designed to achieve.*

In taking into account the *situational* context of the communication being analyzed the investigator considers *who* is speaking, *to whom*, and *under what circumstances.* Clues

to the speaker’s intended meanings [are obtained] by considering various known characteristics of the speaker, his audience, and the nature of the speaker-audience relationship. The investigator also takes into account the time and place of the communication and related events preceding or accompanying it. He does so in the expectation that the exact intended meanings of the words employed by the speaker are shaped by (and understood by the audience with reference to) certain aspects of the setting and the stream of related events. This is particularly likely when, as in wartime propaganda, the communication being analyzed is highly situation- or event-oriented.

Such analysis of the instrumental aspect of communications in their situational contexts is not confined to non-frequency approaches, for it is by no means the case that quantitative or frequency analysis always limits itself to coding “manifest” content. In fact, the criterion of “manifest” content is not generally accepted as essential to the technique of quantitative content analysis. Cartwright (1953) explicitly rejects it, for example. Even in the area of political communication research, for which Lasswell’s version of content analysis has been primarily developed, many if not most quantitative content analyses do not in practice employ the “manifest” content criterion. Rather, they often attempt to infer intended meanings and employ relatively flexible and interpretative procedures for coding content. If this is a drawback from a scientific standpoint, it is one present in many quantitative studies as well as in non-frequency analyses. The difference in this respect between frequency and non-frequency analyses, therefore, seems to be one of degree, stemming partly from the different nature of the two approaches. The non-frequency approach, by virtue of the limited number of cases with which it deals, requires that all relevant intended meanings always be

\*Berelson’s (1952:14–18) definition of content analysis restricts the technique to the analysis of the manifest content of communications, commonly equated with the existence of widespread agreement on what a communication means. Since agreement is also equated with reliability, Berelson concludes that only manifest content can be analyzed reliably. He thus confounds the methodological requirement of reliability with the conceptual distinction between manifest and latent content. Experts may well achieve high reliability in coding latent meanings.

estimated as validly as possible and, therefore, that full account be taken of situational and behavioral contexts.

The emphasis on validity in the non-frequency approach is accompanied by less concern with the reliability of the judgments, or inferences, being made of the speaker's intended meanings. Rarely are systematic procedures employed to ensure or to demonstrate the reliability of non-frequency content descriptions. A possible explanation for this may be suggested. Since the non-frequency analyst works with a relatively small amount of content data, which he collects himself, he tends to be less self-conscious about the reliability problem and less concerned with it than the quantitative analyst is. As a result, he is likely to overlook what the well-trained quantitative investigator knows so well and rightly emphasizes, namely, that when the procedure for obtaining content data on intended meanings is highly interpretative, it is all the more necessary to assess in some fashion the reliability of its results.

#### CLOSE RELATIONSHIP BETWEEN DESCRIPTIVE AND INFERENCE PROCEDURES

Some possible circularities of procedure beleaguer the non-frequency approach to the analysis of communications content. . . . [C]ontent description is intimately intertwined with and overlaps the assessment of inferences from the contents. Inferences as to what the propagandist is trying to say and why he is trying to say it are not neatly discrete.

To illustrate, if one person addresses another as "you old rascal," the analyst who is seeking to interpret the intent validly will want to know if the addressee is an old man or an infant. If it is a baby, one infers that the intent is affectionate and simultaneously describes the content as endearment. There is a mutually interdependent set of assumptions here. One has not established the intent independently and derived the content interpretation from that nor has one established

the affectionate meaning of the phrase "you old rascal" independently and derived the intent from that. The two propositions are parts of an interdependent set of inferential hypotheses.

The question arises of whether this aspect of the non-frequency approach necessarily entails the danger of analytical bias, or "circularity." That is, by not distinguishing more sharply—as in quantitative content analysis—between the descriptive and inferential phases of research, does not the investigator risk the possibility that a hypothesis formulated early in the course of his content description will determine what he subsequently "sees" and regards as significant in the communication?

The danger of circularity in this sense is indeed potentially present in the procedure described above and undoubtedly occurs in many low-grade analyses. However, the disciplined analyst guards against it in several ways. He does not read through the communication material just once, but rereads it as many times as necessary to satisfy himself that the inference favored by him is consonant with all of the relevant portions and characteristics of the original communication material. Similarly, he considers not just one inferential hypothesis when reading and rereading the original communication material, but also many alternatives to it.

He systematically weighs the evidence available for and against each of these alternative inferences. Thus, the results of his analysis, if fully explicated, state not merely 1) the favored inference and the content "evidence" for it, but also 2) alternative explanations of that content "evidence," 3) other content "evidence" which may support alternative inferences, and 4) reasons for considering one inferential hypothesis more plausible than others.

In this fashion, the disciplined analyst controls the dangers of circularity present in the overlapping of descriptive and inferential procedures. To the extent that he operates in the systematic, disciplined fashion we have outlined, the non-frequency analyst follows the accepted scientific procedure of successive approximations.

## NOTES

1. For a brief exposition of quantitative content analysis and some of its uses in the study of political communication, see Lasswell, Lerner, and Pool (1952). These authors' sober assessment of the difficulty of meeting various prerequisites of statistical content analysis is particularly useful.

2. On the difference between interpretation of "intent" and interpretation of "symptoms" in the analysis of communication, see Kecskemeti (1952:61–62).

3. For an explicit statement of this important (but often ignored) assumption of quantitative content analysis, see Berelson (1952:20).

4. This point is explicitly discussed in Goldstein (1942:26–27, 38–40, 150).

5. For a detailed discussion, see Lazarsfeld and Barton (1951:155–192). See also Cartwright (1953).

6. The problem of validity noted here arises only when "manifest" content is used as a rule-of-thumb substitute for intended meaning. It does not arise, of course, when the investigator is interested *only* in the usual meanings of words, as in linguistic studies or in studies of effect on a mass audience rather than of intent of the communicator. In other words, depending upon the hypotheses and questions, which are being investigated, the analyst may be interested either in "manifest" meaning or intended meaning. And, when interested in intended meaning, it can be inferred directly in

each instance or employ "manifest" meaning as a rough approximation.

## REFERENCES

- Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: Free Press.
- Cartwright, D. P. (1953). Analysis of qualitative material. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences* (pp. 421–470). New York: Holt, Rinehart & Winston.
- Goldstein, J. (1942). *Content analysis: A propaganda and opinion study*. Unpublished master's thesis, New School for Social Research, New York.
- Kecskemeti, P. (1952). *Meaning, communication and value*. Chicago: University of Chicago Press.
- Lasswell, H. D., Lerner, D., & Pool, I. de Sola. (1952). *The comparative study of symbols*. Stanford, CA: Hoover Institute Studies, Ser. C: Symbols.
- Lazarsfeld, P. F., & Barton, A. H. (1951). Qualitative measurement in the social sciences: Classification, typologies and indices. In D. Lerner & H. D. Lasswell (Eds.), *The policy sciences: Recent developments in scope and method* (pp. 155–192). Stanford, CA: Stanford University Press.
- Lochner, L. (Ed.). (1948). *The Goebbels diaries*. Garden City, NY: Doubleday.



# 3.5

## EVALUATIVE ASSERTION ANALYSIS

OLE R. HOLSTI\*

### INTRODUCTION

Content analysis—which may be performed using many different techniques, depending upon the theoretical interests of the investigator—is used as a tool for research in international conflict on the premise that from . . . the decision-makers' messages, valid inference may be drawn concerning the attitudes of the speaker or writer. [A] method of content analysis must fulfill a number of requirements.

1. It must provide valid results.
2. It must provide reliable results.
3. It must provide results that are capable of quantification. A continuing study of world tension levels, for example, requires a technique that provides not only a measure of the appearance or non-appearance of certain attitudes, but also of the intensity of those attitudes.

One method meeting these requirements is “evaluative assertion analysis,”<sup>1</sup> a form of quantitative content analysis in which messages are translated into simple, three-element assertive format. Numerical values are then assigned to the constituent elements of each assertion, depending upon its direction and intensity.

Evaluative assertion analysis is not merely a technique for scaling previously coded data. Rather, it is an all-inclusive method of content analysis; as such, it prescribes comprehensive rules for each step from the initial preparation of the written text through the final analysis of the processed data.

This technique was designed for the study of evaluative attitudes on a “good-bad” continuum; its senior author has demonstrated elsewhere, through factor analysis of semantic differentials, that the good-bad, active-passive, and strong-weak dimensions dominate human expression (Osgood, Suci, & Tannenbaum, 1957:50–51, 72–73). The mechanics of the method,

\*From Holsti, O. R. (1963). Evaluative assertion analysis. In R. C. North, O. R. Holsti, M. G. Zaninovich, & D. Zinnes (Eds.), *Content analysis: A handbook with applications for the study of international crisis* (pp. 91–102). Evanston, IL: Northwestern University Press.

however, are suitable for analysis of any dimension defined as a continuum between polar opposites. The technique is also readily adaptable for measuring categories that are defined, as in Q-Sort scaling, as a single “more-to-less” continuum. For the study of international conflict, relevant variables, in addition to those mentioned above, include: hostility-friendship, satisfaction-frustration, strength-weakness, specificity-diffuseness, and violence-nonviolence. Any dimensions chosen for analysis must, of course, be explicitly defined.

As with all kinds of content analyses, evaluative assertion analysis rests upon certain minimal premises regarding (1) the structure of messages, and (2) the operations that can be undertaken by reasonably skilled coders with an acceptable degree of reliability. These assumptions have, however, been empirically shown to be valid (Osgood, Saporita, & Nunnally, 1956:47–48).

It will suffice [here] . . . to outline very briefly the primary characteristics of this technique. A reader contemplating the use of evaluative assertion analysis should turn to the original source for a comprehensive description. The summary below will serve as an introduction to the method and as the basis for discussing its utility in research on international conflict (Osgood, 1959; Osgood et al., 1956).

## CODING AND SCALING

The steps for converting unedited messages into the quantified data against which hypotheses can be tested are as follows:

1. The initial step in evaluative assertion analysis is the identification and isolation of

attitude objects in relation to the variables under study (Osgood et al., 1956:49). Attitude objects are symbols whose evaluative meanings vary from person to person; for example, capitalism, foreign aid, United Nations, Khrushchev. Common-meaning terms are those whose evaluative meanings vary, minimally; for example, evil, honest, benevolent. In general, terms that are capitalized are attitude objects rather than common-meaning terms.

2. After attitude objects—which might include nations, policies, ideologies, decision-makers, non-national organizations or general symbols—have been identified by the coders, they are masked with meaningless symbols. For example, the text of a Soviet note to the United States Government states that,

In recent days, fascistic elements with the obvious connivance of the United States occupation authorities have carried out in the American sector of West Berlin a series of dangerous provocations against members of the honor guard of the Soviet forces.

After masking of attitude objects with nonsense symbols, the edited text would read as follows:

In recent days, fascistic elements with the obvious connivance of the AX occupation authorities have carried out in the AX sector of BY a series of dangerous provocations against members of the honor guard of the CZ forces.

Note that because in the Soviet note the terms “United States” and “American” are interchangeable, both are masked with the same symbol.

3. Following these initial operations, the masked message is translated into one of two generic assertion forms:

Form A: Attitude Object<sub>1</sub> (A0<sub>1</sub>)/verbal connector (c)/common meaning term (cm)

Form B: Attitude Object<sub>1</sub> (A0<sub>1</sub>)/verbal connector (c)/Attitude Object<sub>2</sub> (A0<sub>2</sub>)

Comprehensive guides for translation of the text have been prepared (Osgood et al., 1956:59–89), making possible the revision of the most complex sentences. Thus an editorial

statement in *Jen-min jih-pao*: “The treacherous American aggressors are abetting the corrupt ruling circles of Japan,” would be coded as follows:

## 158 • PART 3

1.	Americans	are	treacherous	(form A)
2.	Americans	are	aggressors	(form A)
3.	Americans	are abetting	Japanese ruling circles	(form B)
4.	Japanese ruling circles	are	corrupt	(form A)

The complete text is typed on a seven-column data chart (Figure 1): Values are then entered in columns 4 and 6 of the data chart. If the project involves the analysis of more than one dimension, assertions should be

kept separate either by adding an additional column on the data chart in which identification of the dimension can be made, or by maintaining a separate data chart for each dimension.

1	2	3	4	5	6	7
Source	AO <sub>1</sub>	c	Value of Column 3	cm or AO <sub>2</sub>	Value of Column 5	Product: Columns 4 × 6

**Figure 1**

4. The next step is to determine the direction or valence and intensity of the attitudes, as expressed in the verbal connector and the common-meaning term. Each of these is rated for both valence (+ or -) and intensity (1, 2 or 3). The direction of the verbal connector depends upon whether the perceived relationship is associative (+) or dissociative (-). The valence of the common-meaning term is determined by whether the expressed attitude lies on the negative or the positive side—however these are defined by the researcher—of the neutral point on the dimensional scale.<sup>2</sup>

Intensities for the verbal connectors and common-meaning terms are also assigned according to a comprehensive set of guides. For example, most unqualified verbs or verbal phrases in the present tense are given a value of  $\pm 3$ ; verbs with auxiliaries are rated  $\pm 2$ ; and, verbs implying only a hypothetical relationship are assigned a value of  $\pm 1$ . Similarly, common-meaning terms are rated 1, 2, or 3, corresponding roughly to the categories “extremely,” “moderately,” and “slightly.” The assigned values are then entered in columns 4 and 6 of the data chart.

The values for attitude objects are first determined for all assertions in form A; only

after the values for attitude objects in assertions of form A have been calculated, can the evaluation for assertions in form B be made. In the previous example, assertions 1, 2, and 4 are of type A, whereas assertion 3 (Americans / are abetting / Japanese ruling circles) is in form B. The numerical value of “Japanese ruling circles” is calculated by every assertion of type A. In assertion 4 it was stated that,

Japanese ruling circles / are / corrupt

From this and other assertions of a similar nature, (AO<sub>1</sub> / c / cm), that might appear in the text, it is possible to calculate the perceived evaluation of “Japanese ruling circles” (in this case a strongly negative one). That value is then inserted into assertion 3; thus because the Americans are closely associated (“are abetting”) with the Japanese ruling circles, the value of “Americans” is a strongly negative one.

The reader may ask, “What if the text is composed entirely of assertions in form B, making it impossible to determine any values?” This can only occur in messages devoid of any adjectives or adjectival phrases. Thus, it

is difficult to imagine an extensive communication in which every assertion is of the  $AO_1 / c / AO_2$  type.

5. The scaling of an attitude object on any dimension is the sum of its evaluation in

assertions of form A and form B. In each case the value is the *product* of the second (verbal connector) and third element (common meaning term [form A] or attitude object<sub>2</sub> [form B]). For example, assertion 4 above would appear as follows on the data chart (Figure 2):

1	2	3	4	5	6	7
Source	$AO_1$	c	Value of Column 3	cm or $AO_2$	Value of Column 5	Product: Columns 4 × 6
<i>Jon-min jih-pao</i>	Japanese ruling circles	are	+3	corrupt	-3	-9

Figure 2

The reason for multiplying the values in columns 4 and 6 is to assure the proper valence or direction of the final evaluation; thus the double negative (X is not bad) assertion will receive the same value as the double positive (X is good) assertion.

The final evaluation of each attitude object is calculated in three steps:

1. All values in column seven for assertions of type A are summed.
2. All values in column seven for assertions of type B are summed.
3. The total of the values derived in steps x and z is then divided by the modular sum of column three.

The final evaluation may be expressed algebraically as,<sup>3</sup>

$$\text{Evaluation } AO_1 = \frac{\sum_{i=1}^n c_i cm_i + \sum_{i=1}^n c_i (AO_2)_i}{\sum |c| cm + \sum |c| AO_2}$$

#### APPLICATIONS OF EVALUATIVE ASSERTION ANALYSIS

The results of the completed analysis may be aggregated in a variety of ways. For some

projects, it might be useful to compare single documents, whereas for others the analyst may be interested in compiling totals for all documents within prescribed time periods. In other cases, it may be desirable to combine results in terms of the senders or recipients of the messages.<sup>4</sup> Such a decision will, of course, be dictated by the nature of the research problem.

A number of objections may be raised against evaluative assertion analysis. In the first place, the method is admittedly time consuming.<sup>5</sup> A second point is that the translation of the text into assertion form leads to some loss in the “flavor” of the original message.

There is some weight in both objections, but the technique has many compensating advantages. By translating all messages into assertion form, much is gained by providing a high degree of uniformity for the judges who must do the scaling. Three major sources of low reliability are (1) the ambiguity of categories, (2) confusion over the perceived roles of various attitude objects within a sentence, and (3) difficulty in assigning numerical values to complex statements. Each of these points will be considered in terms of evaluative assertion analysis.

The first problem is primarily a theoretical one and precedes the coding stage. However, a technique which reduces each sentence to its

## 160 • PART 3

constituent elements eliminates the possibility of more than one dimension appearing in any one assertion. This may be illustrated by a typical Chinese statement during the U-2 crisis: "The Chinese people firmly support the stand of the Soviet Government in opposing United States imperialism's war provocation and its sabotage of the Summit Conference." This sentence consists of a number of attitude objects in

a complex relationship. In addition, the sentence contains elements of friendship (firmly support), hostility (oppose, war provocation, sabotage), evaluation (just stand), and policy conditions (firmly support, war provocation, sabotage, opposing). The unedited text clearly poses a problem for the scaler; when coded in assertion form, which separates the various elements, the difficulties are materially reduced.

Chinese people	/	firmly support	/	Soviet Government
Soviet Government's stand	/	is	/	just
Soviet Government	/	opposes	/	the United States
United States	/	is	/	imperialistic
United States	/	provokes	/	war
United States	/	sabotaged	/	Summit Conference

A second source of difficulty with many techniques, arising usually after a sentence has been masked, is the possibility of confusing the perceived roles of the various attitude objects in any sentence. In the statement cited above, for example, there are three actors—the Chinese people, the Soviet Government, and the United States—and maintaining their perceived relationship is of crucial importance. The translation of statements into assertion form minimizes the possibility of confusion because the position of each element in the assertion is always the same. The data sheets themselves impose a high degree of uniformity, being divided into columns, which maintain that order throughout.

As stated elsewhere . . . , the essential theoretical components of any statement are (1) perceiver, (2) perceived, (3) action, and (4) target. Evaluative assertion analysis is readily adaptable to such a conceptualization:

Perceiver	=	Source
Perceived	=	Attitude Object <sub>1</sub>
Action	=	Verbal Connector
Target	=	Attitude Object <sub>2</sub>

In addition, there is a vital fifth element, the incorporated modifiers, which may be connected to the perceiver, perceived, or target. One of the valuable characteristics of evaluative assertion analysis is that it forces a separation, for the purposes of analysis, of "action assertions" from "evaluative assertions."<sup>6</sup> The importance of this point can be illustrated in the statement, "The valiant X has repelled the treacherous forces of Y." Although it includes only one perceiver (author of the statement), one perceived (X), one action (has repelled), and one target (Y), the statement creates difficulties—both for the coder who must categorize it and for the scaler who must assign it a numerical value—owing to the presence of the affective elements "valiant" and "treacherous," in addition to the action element of "has repelled." But when the sentence is translated into

1.	X	/	has repelled	/	Y	(action assertion)
2.	X	/	is	/	valiant	(evaluative assertion)
3.	Y	/	is	/	treacherous	(evaluative assertion),

much of the difficulty, both in categorization and in assigning numerical values, is resolved. Assertion 1 can then be scaled for action dimensions such as activity-passivity, specificity-diffuseness or violence-non-violence; assertions 2 and 3 can be scaled for affective dimensions such as good-bad or hostility-friendship.

When this fifth element, the incorporated modifier, has been introduced as a separate constituent, the conversion between the theoretical framework developed in this manual and evaluative assertion analysis is complete:

Perceiver	=	Source
Perceived	=	Attitude Object <sub>1</sub>
Action (or attributive verb)	=	Verbal Connector
Target	=	Attitude Object <sub>2</sub>
Incorporated Modifiers	=	Common-meaning terms

A third source of low reliability—difficulty over the assignment of numerical values to complex sentences—is reduced to a minimum by allowing the scaler to focus attention first on the verbal connector and then on the common-meaning term, in each case a single word or a short phrase. When all data have been processed, it is possible to do a rapid congruity check on the finished data sheets to detect any errors (Osgood et al., 1956:98–99).

Unlike forced distribution scaling techniques, evaluative assertion analysis is amenable to comparative analysis *across* as well as *within* universes of statements. For example, a project may involve scaling all Soviet statements in the month before the U-2 incident and the month after the affair as separate bodies of data, in order to test hypotheses concerning the patterns of variables. If, however, it is also desirable to compare hostility levels between the two months, this cannot be done using any forced distribution scaling technique without further rescaling of at least samples from the combined universes, because the mean hostility level for

each month is by definition identical.<sup>7</sup> While this additional step is by no means an insurmountable barrier, a technique which defines the value of each value category rather rigorously beforehand bypasses some of the problems of comparative analysis.

#### ADAPTABILITY TO COMPUTER ANALYSIS

A final point, which may be considered, is the adaptability of evaluative assertion analysis to computer analysis. Translation into assertion form appears to be one of the methods most readily adaptable to this type of analysis.\* Retrieval of relevant assertions, assignment of values, and the arithmetic computations can easily be performed by computer. Finally, the results can be aggregated in terms of the researcher's hypotheses.

#### CONCLUSION

It is almost inevitable that research into international conflict involving any extensive use of content analysis will be group research, utilizing teams of translators, coders, scalers, data recorders, programmers, analysts, and others. Because both coders and scalers are likely to be part-time and short-term employees of the research project, the rules for coding and scaling must be sufficiently comprehensive to avoid ambiguity, yet simple enough to be easily learned. For this reason, a technique of content analysis, such as evaluative assertion analysis—by imposing a high degree of uniformity on each of the various steps of data preparation and analysis—can be of great value. Moreover, research personnel can be rapidly trained. The increment of additional time required to use evaluative assertion analysis must be weighed against the degree of reliability and precision that is gained; in the end, however, the selection of a methodological tool must rest upon the nature of the research problem and the information

\*See Kleinnijenhuis, de Ridder, and Rietberg (reading 7.5, this volume), who use computer aids to kernelize text in the above two and several additional types of assertions.

that the researcher seeks to obtain from the communications to be analyzed.

## NOTES

1. The most complete guide to this technique is Osgood et al. (1956). A briefer, and more readily accessible, summary may be found in Osgood (1959:41–54). The present brief description of the various steps in evaluative assertion analysis is derived from these sources.

2. The continuum has a middle point of zero. Any one statement, however, with an evaluative product of zero should not be coded. For example, the assertion “Kennedy is a man,” has a value of zero on a friendship-hostility scale. Thus the statement is not coded.

3. This formula gives a “weighted” evaluation (see Osgood et al., 1956:92). For research of this kind described in this manual, an un-weighted evaluation (in which each assertion is given equal value), may be more desirable. In this case, the following formula may be used:

$$\text{Evaluation } AO_1 = \frac{\sum_{i=1}^n c_i cm_i}{3^n} + \frac{\sum_{i=1}^n c_i (AO_2)_i}{3^n}$$

in either case, final evaluations fall within a range of +3 to –3. The rationale for using the un-weighted evaluation formula is discussed in Appendix A of Holsti (1962).

4. Examples of the various uses of evaluative assertion analysis may be found in Holsti (1962a) and (1962b).

5. Coders are able to process completely about one page per hour. A short form of the method is described in Osgood et al. (1956:96–97). Coding

speed can be increased by a factor of three without a disastrous loss of inter-coder reliability.

6. Adjectives formed from verbs or implying an object may cause some ambiguity. Consider the assertion, “X is aggressive.” This is both evaluative on a number of scales (hostility, friendship, etc.) and implies action against an unspecified or general target. This point was raised by William Quandt (1962).

7. It should be noted that there are dangers inherent in the assumptions that all results are comparable. For example, even a cursory reading of Chinese Communist statements will reveal a level of affect rarely found in the more genteel diplomatic language of the nineteenth century.

## REFERENCES

- Holsti, O. R. (1962a). *The belief system and national images: John Foster Dulles and the Soviet Union*. Ph.D. dissertation, Stanford University.
- Holsti, O. R. (1962b). The belief system and national images: A case study. *Journal of Conflict Resolution* 6:245–252.
- Osgood, C. F. (1959). The representational model. In I. de Sola Pool (Ed.), *Trends in content analysis* (pp. 33–88). Urbana: University of Illinois Press.
- Osgood, C. F., Saporta, S., & Nunnally, J. C. (1956). Evaluative assertion analysis. *Litera* 3:47–102.
- Osgood, C. F., Suci, G. J., & Tannenbaum, P. I. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Quandt, W. (1962). The application of the General Inquirer to content analysis of diplomatic documents. *Stanford Studies in International Conflict and Integration*, July. 1.

# 3.6

## AN ECOLOGY OF TEXT

### *Memes, Competition, and Niche Behavior*

MICHAEL L. BEST\*

#### INTRODUCTION

Ideas do not exist in a vacuum. Neither does discourse, the interconnected ideas that make up conversation and texts. In this research, we investigate the pair-wise interaction between populations of ideas within discourse: Are our text populations in competition with each other? Do they mutually benefit each other? Do they prey on one another?

This work attempts to build models of *population memetics* by bringing together two disciplines: Alife\*\* and text analysis. Through techniques of text analysis, we determine the salient co-occurring word sets, texts, and text clusters, and track their temporal dynamics. We then study the life-like properties of this human-made system by considering its behavior in terms of replicators, organisms, and species.

Richard Dawkins coined the term *meme* to describe replicating conceptual units (Dawkins, 1976). In studying the population

dynamics of ideas we consider the meme to be the largest reliably replicating unit within our text corpus (Pocklington, 1996; Pocklington & Best, 1997). Through text analysis, we identify memes within a corpus and cluster together those texts, which make use of a common set of memes. These clusters describe species-like relationships among the texts.

The particular texts we study are posts to the popular USENET News (or NetNews) system. These posts form the basis of a new Alife environment, the *corporal ecology* (Best, 1996, 1997). In this ecology, texts are the organisms, the digital system defined by NetNews describes an environment, and human authors operating within some culturally defined parameters are the scarce resource.

At the core of our study sits a large text analysis software system based primarily on Latent Semantic Indexing (LSI) (Deerwester, Dumais, Furnas, Landauer, & Harshman,

\*Best, M. L. (1997). Models for interacting populations of memes: Competition and niche behavior. *Journal of Memetics—Evolutionary Models of Information Transmission, 1*.

\*\*Short for artificial life, the study of computer simulations of living systems and their evolution.



1990; Dumais, 1992, 1993; Furnas et al., 1988). This system reads each post and computes the frequency with which each word appears. These word counts are then used in computing a vector representation for each text. A principal component analysis is performed on this collection of vectors to discover re-occurring word sets; these are our memes. Each post is then re-represented in terms of these memes. By grouping texts, which are close to one another within this meme-space, we cluster semantically similar texts into species-like categories or *quasi-species* (Eigen, McCaskill, & Schuster, 1988).

We proceed to study the interactions between those populations that coincide temporally. For each cluster, we compute a series that represents its volume of post activity over time, for instance, how many texts of a given cluster were posted on a given day. Cross-correlations between each pair of time series are then determined. We find that some pairs have strong negative correlations and argue that these are examples of texts in competition. A number of examples of such competition are explored in depth. We argue that high competition is correlated with those text clusters that exist within a narrow ecological niche; this phenomenon is also observed in natural ecologies (Pianka, 1981).

Note that this is an unusual shift from the typical Alife environment. We are not synthesizing replicators, embodying them into agents, and observing their life-like interactions. Instead, we are studying a pre-existing artifact. Through our analysis, we *discover* replicators within organisms, and use computational techniques to observe their dynamics.

In this paper we first briefly overview the NetNews environment and describe the LSI-based text analysis system. Next, we describe the mechanism used to determine the temporal dynamics and cross-correlations given a corpus of posts. We then relate the cross-correlations to models of interacting populations. In the next section, we examine in depth a couple pairs of post clusters with strong interactions. We then describe a theory of niches within the corporal ecology and note that narrow ecological niches are

correlated with significant competition. We end with our conclusions.

## THE NETNEWS CORPUS

---

Understanding our corpus requires a basic knowledge of the NetNews system. NetNews is an electronic discussion system developed for and supported on the Internet (Kantor & Lapsley, 1986). Discussion groups have formed along subjects ranging from science to politics to literature to various hobbies. The collections of messages are organized into particular subject groups called *newsgroups*. The newsgroups themselves are organized in a tree-like hierarchy, which has general top-level categories at the root and moves to more specific topics as you progress towards the leaves. A newsgroup name is defined as the entire path from the top-level category through any subsequent refining categories down to the name of the group itself. Category and group names are delimited by the period symbol. Thus, "soc.religion" is the name of a newsgroup concerned with social issues around the world's religions and "soc.religion.hindu" is a more specific group devoted to Hinduism.

Texts sent to NetNews, the *posts*, are composed of a number of fields only a few of which are relevant here. The user creating the post is responsible for the post body (that is, the actual text of the message) as well as a subject line. The subject line is composed of a few words that describe what the post is about. NetNews software will attach a number of additional fields to posted messages including a timestamp and the user name of the person who created the post.

Posts can be either an independent message or a follow-up to a previous message. A follow-up, or "in-reply-to" message, will have special threading information in its header linking it to the previous posts to which it is a reply. This header information allows newsreaders to reconstruct the discussion thread.

NetNews today has grown considerably from its beginnings in the late 70's and 80's. With over 80,000 posts arriving each day, it provides an excellent dataset for the study of cultural microevolution.

## THE TEXT ANALYSIS METHOD

We analyze a corpus of posts to NetNews to distill their salient replicating unit or memes, and to cluster together posts, which make common use of those memes. We do this by employing a large system of text analysis software we have built. The techniques employed are based on the vector space model of text retrieval and Latent Semantic Indexing (LSI).

### Vector Space Representation

We begin with a corpus composed of the full-text of a group of posts. We analyze the corpus and identify a high-dimensional space, which describes the conceptual elements within the texts. For each post, we identify a point within this space, which captures it semantically. This technique is known as a vector space representation (Frakes & Baeza-Yates, 1992; Salton & Buckley, 1988). Each dimension in this space will represent a *term* from the corpus where a term is a word that occurs with some frequency (e.g., in at least three posts) but not too frequently (e.g., the word “not” is dropped from the term list). The goal is to arrive at a set of terms that semantically capture the texts within the corpus.

Given the conceptual space described by this set of terms, each post can be represented as a point within this space. We score each document according to the frequency each term occurs within its text, and assign each term/document pairing this *term weight*. The weighting we use for each term/document pair is a function of the *term frequency* (simply the number of times the term occurs in the post) and the *inverse document frequency* (IDF). Consider a corpus of  $m$  posts and a particular term,  $j$ , within a list of  $n$  terms. Then the IDF is given by,

$$IDF_j = \log\left(\frac{m - m_j}{m_j}\right),$$

where  $m_j$  is the number of posts across the entire corpus in which term  $j$  appears. Thus, if a term occurs in 50% or more of the texts the IDF for that term will vanish to zero. But if,

for instance, a term occurs in 10% of the documents the IDF will be nearly  $\log(10)$ . In words, rare terms have a large IDF.

The term weight for a document,  $i$ , and term,  $j$ , is then defined by,

$$TermWeight_{ij} = w_{ij} = \log(TermFrequency_{ij}) \cdot IDF_j.$$

Each term weight, then, is a function of the inter- and intra-document term frequencies.

Each post,  $i$ , is now represented by a particular term vector,  $r_i = (w_{i1}, w_{i2}, \dots, w_{in})$ . The entire collection of  $m$  term vectors, one for each post, define the *term/document matrix*,  $A$ .

This set of steps, culminating in the term/document matrix, forms the basis for much of modern text retrieval or filtering and is at the core of most Web search engines.

### Latent Semantic Indexing (LSI)

LSI is a technique used to distill high-order structures from a term/document matrix, consisting of sets of terms that re-occur together through the corpus with appreciable frequency. The re-occurring term sets are discovered through a principal component method called Singular Value Decomposition (SVD). While LSI was primarily developed to improve text retrieval, we are interested in its ability to find replicating term sets, which act as memes. We will first overview the LSI technique and then discuss how it discovers memes.

LSI was originally proposed and has been extensively studied by Susan Dumais of Bell Communications Research and her colleagues (Deerwester et al., 1990; Dumais, 1992, 1993; Furnas et al., 1988). Peter Foltz investigated the use of LSI in clustering NetNews articles for information filtering (Foltz, 1990). Michael Berry and co-authors researched a variety of numerical approaches to efficiently perform SVD on large sparse matrices such as those found in text retrieval (Berry, 1992; Berry & Fierro, 1995; Berry, O'Brien, Do, Krishna, & Varadhan, 1993).

The SVD technique decomposes the term/document matrix into a left and right orthonormal matrix of eigenvectors and a

diagonal matrix of eigenvalues.\* The decomposition is formalized as,  $A_k = UV^T$ .

The term/document matrix,  $A$ , is approximated by a rank- $k$  decomposition,  $A_k$ ; in fact the SVD technique is known to produce the *best* rank- $k$  approximation to a low-rank matrix (Berry, 1992).

We are interested in only the right orthonormal matrix of eigenvectors,  $V^T$ . Each row of this matrix defines a set of terms whose co-occurrences have some statistically salient re-occurrences throughout the corpus. That is, each eigenvector describes a subspace of the term vector space for which the terms are frequently found together. These *term-subspaces* describe a set of semantically significant associative patterns in the words of the underlying corpus of documents; we can think of each subspace as a *conceptual index* into the corpus (Furnas et al., 1988).

For instance, an example term-subspace generated by analyzing a collection of military posts found three words as having significant re-occurrences, and therefore replicating together with success: “harbor,” “japan,” and “pearl.” These term-subspaces make up our replicators and are our putative memes. Memes are not single re-occurring words but are made up of *sets* of re-occurring words.

Our final text analysis step is to “compress” the original term/document matrix by multiplying it with this right orthonormal matrix of eigenvectors (in other words we perform a projection). This, in effect, produces a *term-subspace/document* matrix. Each post is represented by a collection of weights where each weight now describes the degree to which a term-subspace is expressed within its post’s text.

## MEME AND QUASI-SPECIES

### Term-Subspace as Putative Meme

We are looking for replicators within the corpus that are subject to natural selection. Elsewhere we have argued at length as to why

the term-subspace captures the requirements of a true meme because its word sets act as a unit of selection within the corpus (Best, 1996, 1997; Pocklington & Best, 1997). The strengths of this term set as a replicating unit of selection are due to it meeting the following conditions:

- it is subject to replication by copying,
- it has strong copying fidelity,
- but not perfect fidelity, it is subject to mutation,
- it has a strong covariance with replicative success (Eigen, 1992; Lewontin, 1970).

We will quickly review each of these points in turn.

SVD techniques exploit structure within the term/document matrix by locating co-occurring sets of terms. Clearly, these term sets are replicating through the corpus since that is the precise statistical phenomena the SVD analysis detects. However, it is not obvious that this replication is generally due to copying. Instances of precise copying occur when an in-reply-to thread includes elements of a previous post’s text via the copying mechanism provided by the software system. Other instances of copying occur within a particular context or discussion thread when authors copy by hand words or phrases from previous posts into their new texts. More abstractly, replication occurs because certain memes are traveling outside of the NetNews environment (and thus outside of our means of analysis) and authors again act as copying agents injecting them into the corporal ecology. But, clearly, some re-occurrences are not due to copying but are a chance process where unrelated texts bring together similar words. The likelihood of such chance re-occurrences will be a function of the size and quality of our replicating unit. In summary, term-subspaces are instances of replication often due to copying.

The copying fidelity of a term-subspace is also a direct outcome of the SVD statistical

\*The eigenvalue of a transformation is that value on which it converges after infinitely many applications of that transformation. The eigenvector of a transformation, here  $T$ , is that vector whose direction remains unchanged by that transformation.

analysis. But importantly, the copying fidelity of re-occurring term sets is not perfect across the entire corpus; the term sets will co-occur with some variation. These mutations are both changes *designed* by human authors and chance variation due to copying errors. In either case, the mutations are *random* from the vantage of selection; in other words, human authors are not able to perfectly predict the adaptive significance of their inputted variations. These mutations work “backwards” into the actual term-subspace representation for a post organism. That is, a random mutation at the post level will actually result in a random mutation in the vector subspace representation (the memotype) for the post organism. In this way, the memes as represented in the memotype are subject to mutation.

Finally, we have elsewhere shown there can be a strong covariance between the replicative success of a cluster or thread of posts and the degree to which they express certain term-subspaces (Pocklington & Best, 1997). In other words, a group of posts can increase its volume of activity over time by increasing the degree to which it expresses certain term sets within its post’s text. This, then, is a covariance between the *fitness* of a population of posts and the expression of a particular *trait* as defined by a term-subspace. The demonstration of this covariance is critical to establishing that a replicator is subject to natural selection.

### Quasi-Species

If the term-subspace is a reasonable model for the meme then the term-subspace vector representation of a post is a good model of the post’s *memotype*. Much as a genotype describes a point within genetic sequence-space for each organism, the memotype describes a point within conceptual sequence-space. By *sequence-space*, we mean any of the search spaces defined by a replicator undergoing selection. Examples of sequence-spaces include the gene space, protein spaces under molecular evolution, and the meme space defined within a corporal ecology.

The notion of a *quasi-species* is due primarily to Manfred Eigen (Eigen, 1992; Eigen

et al., 1988). He states that the “quasi-species represents a weighted distribution of mutants centered around one or several master sequences. It is the target of selection in a system of replicating individuals that replicate without co-operating with one another (RNA molecules, viruses, bacteria)” (Eigen, 1992). One organism is a *mutant* of another if it is particularly close to the other in sequence-space.

We wish to group our posts into quasi-species. This requires finding groups of memotypes that are centered together within the conceptual sequence space. To do so we employ a simple clustering algorithm, the Nearest Neighbor Algorithm (Jain & Dubes, 1988). We first normalize each post memotype to unit length; this amounts to discarding text length information and representing only the *relative* strength of each meme within a text. The clustering algorithm then considers each post memotype in turn. The current memotype is compared to each memotype, which has already been assigned to a cluster. If the closest of such vectors is not farther than a threshold distance, then the current vector is assigned to that cluster. Otherwise, the current vector is assigned to a new cluster. This continues until each and every vector is assigned to a cluster.

This process assigns each post to a quasi-species defined as those posts which are close to one another in conceptual sequence-space.

The overall aim in grouping organisms is to bring to light certain evolutionarily significant relationships. Clearly, our quasi-species clustering method is a-historical; that is, it does not directly account for descent when grouping together text organisms. The extent to which such groupings are effective when studying the relatedness of natural organisms is a matter of continued controversy as can be seen in the debates of the cladists versus evolutionary systematists versus pheneticists. While we are currently agnostic to this controversy, we do agree with an original claim of the pheneticists: the more traits are used when assessing the relatedness of individuals the more accurate are the groupings (Mettler, Gregg, & Schaffer, 1988).

We are in the happy situation of clustering based on the complete memotype for each of our organisms. The result is that under

empirical verification our clusters exhibit extremely strong historical relatedness. We have found that the vast majority of texts clustered together come from the same in-reply-to thread and thus are related by descent (Best, 1997). But our clustering method has the added benefit of grouping related texts even when the in-reply-to mechanism is not used and, alternatively, breaking up texts that are within the same thread but are not semantically related. This is of value since many posters to NetNews use the in-reply-to mechanism to post unrelated texts or, alternatively, post follow-up texts without bothering to use the in-reply-to facility. Thus, we claim that our clustering mechanism, due to its access to hundreds of traits, is actually superior at grouping together both related and descendent texts then would be a simple reliance on the threading mechanism. The clustering method meets our goal of illuminating evolutionarily significant relationships.

### Comparison to Natural Ecologies

We are describing phenomena within a corpus of texts in terms of population ecology and population genetics. This is not simply a metaphorical device; we believe that interacting populations of texts and their constituent memes are evolving ecologies quite exactly. However, there are clearly a number of interesting differences between genes and memes (as here operationally defined), natural organisms and texts, ecologies and corpora. Important differences include the driving forces behind mutation within the texts and the role of self-replication and lineage within the corpora. We leave to future work a more complete analysis of these differences.

### MODELS FOR INTERACTING POPULATIONS

We now turn to studying the interaction between quasi-species of posts. We have so far only studied the pair-wise interactions between post quasi-species. Similar pair-wise interactions have been widely studied within

theoretical ecology. Consider two interacting populations: one population can either have a positive effect (+) on another by increasing the other's chance for survival and reproduction, a negative effect (–) by decreasing the other population's survival chances, or a neutral (0) effect. The ecological community has assigned terms to the most prevalent forms of pair-wise interaction, in particular:

- Mutualism (+, +)
- Competition (–, –)
- Neutralism (0, 0)
- Predator/prey (+, –)

(May, 1981; Pielou, 1969).

Our goal is to study the pair-wise interactions of quasi-species within the corporal ecology with the hope of discovering some of these interaction types.

### Time Series

To study how the interactions of populations affect growth rates we must define a method to measure a quasi-species' growth over time. Recall that a quasi-species describes a collection of posts, which are close to one another in sequence-space. Each of these posts has associated with it a *timestamp* identifying when that text was posted to the system; in effect, its birth time and date. (Note that a post organism has something of a zero-length life-span; it comes into existence when posted but has no clear time of death.)

A histogram of the timestamp data is created with a 24-hour bucket size. That is, for each quasi-species we count how many member texts were posted on one day, how many on the next, and so forth through the entire population of texts. The datasets currently used span on the order of two weeks and consist of thousands of posts. So, for each day a quasi-species has a volume of activity, which can range from zero to 10's of posts. This rather coarse unit, the day, has been chosen to neutralize the strong daily patterns of post activities (e.g. activity may concentrate in the afternoons and drop off late at night, different time zones will shift this behavior

and thus encode geographic biases). Thus, the patterns of rise and fall in the volume of posts within a quasi-species when measured at the day level will, hopefully, reflect true changes in interest level and authorship activity rather than other external or systemic factors.

### The Test Corpus

Figure 1 is a typical graph for the volume of posts within a particular quasi-species over a period of ten days. This cluster was found within a corpus of all posts sent to the soc.women newsgroup between January 8, 1997 (the far left of the graph) and January 28, 1997 (the far right). In the figure, the number of posts in a day is represented by the height of the graph. This particular cluster of texts exhibited an initial set of posts, a few days worth of silence, and then a rapid building up of activity that finally declined precipitously at the end of the dataset. The entire corpus used consisted of 1,793 posts over the same ten day period. The clustering mechanism arrived at 292 quasi-species, the largest of which contained 103 posts.

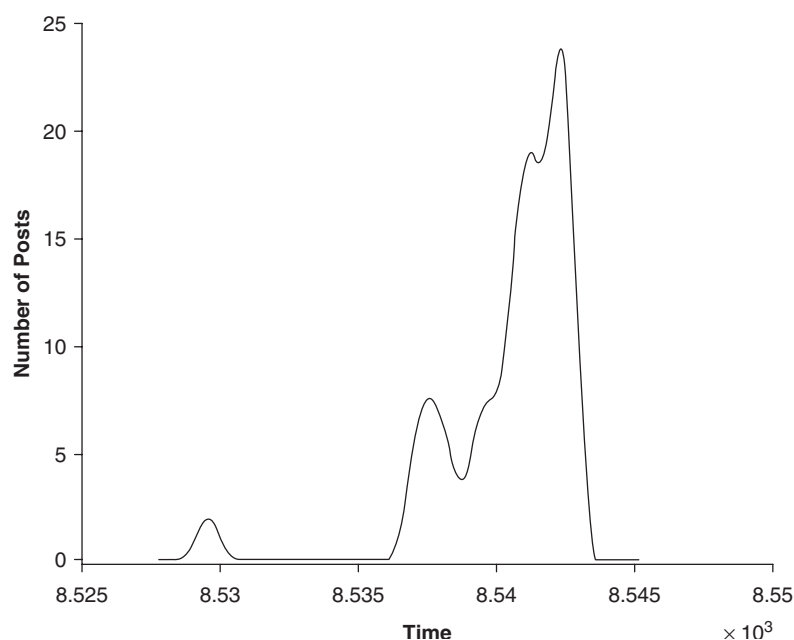
### Time Series Cross-Correlation

To study the relationship between the time series of two populations of posts we use the cross-correlation function. The use of the cross-correlation to study bivariate processes, and time series in particular, is well known (Chatfield, 1989). Each time series is normalized to be of zero mean and unit standard deviation; that is, we subtract off the mean and divide by the standard deviation. In this way, the cross-correlations will not be dominated by the absolute volume of post activity within some cluster and instead will be sensitive to both large and small sized clusters.

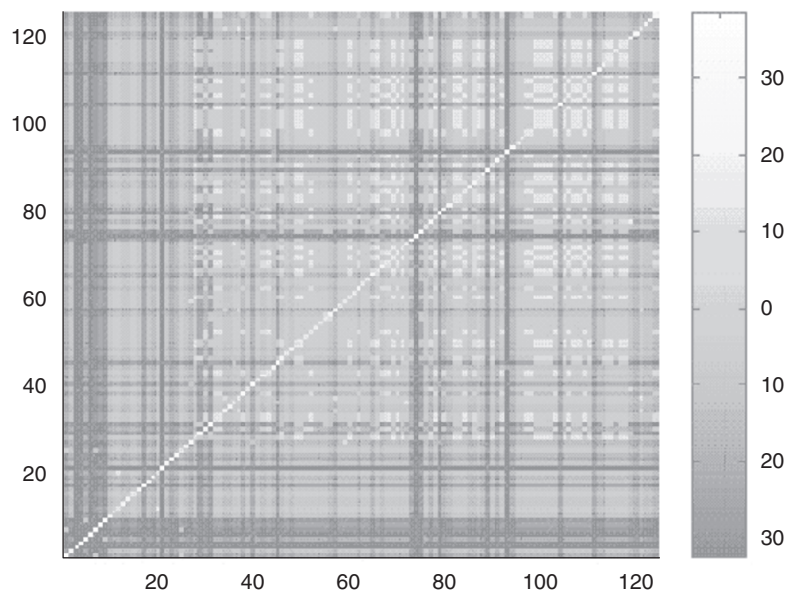
We assume the readers are familiar with the regular covariance and correlation functions. Then the cross-correlation for two time series,  $X$  and  $Y$ , is given by

$$\rho_{xy} = \frac{\gamma_{xy}}{\sqrt{\gamma_{xx}\gamma_{yy}}}$$

Here,  $\gamma_{xy} = \text{Cov}(X, Y)$  and  $\gamma_{xx}$  and  $\gamma_{yy}$  are the variance of  $X$  and  $Y$  respectively. Note this



**Figure 1** A typical time series of posts to quasi-species. Time axis is measured in seconds since Jan. 1 1970.



**Figure 2** The Pair-Wise Time Series Cross-Correlation for 125 Largest Quasi-Species Clusters

formulation only considers the cross-correlation for a zero time lag. That is, it considers how the two time series are correlated for identically matching points in time. With a nonzero lag the cross-correlation would study cases when the two series might have correlations offset by some fixed amount of time. Since we group our time data into day-long chunks the zero-lag cross-correlation will be sensitive to covariances, which have a time offset as large as 24 hours; this builds into the time series an adequate time lag.

When the cross-correlation between two sets of data is significantly different than zero it suggests the two sets of data have some relationship between them. A positive value means an increase in one series is likely to co-occur with an increase in the other series. A negative value means an increase in one series is likely to co-occur with a decrease in the other series.

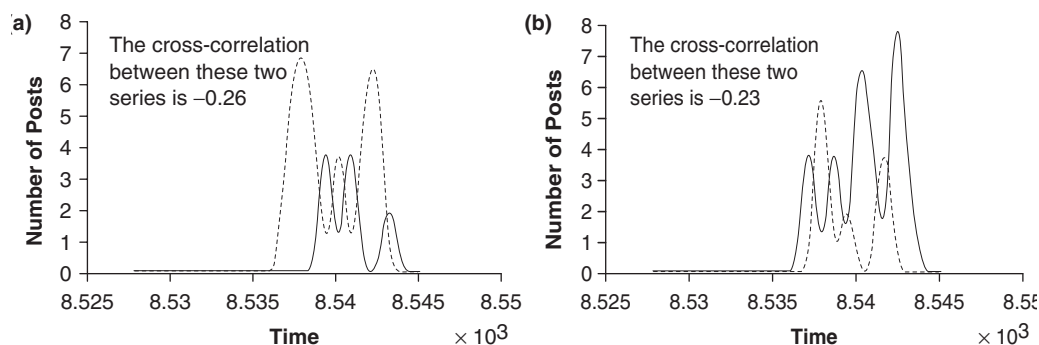
Figure 2 shows the pair-wise cross-correlations for the 125 largest quasi-species clusters within our corpus. The diagonal represents the cross-correlation between a time series and itself which, as expected, is identically one. Note that the matrix is symmetric

about the diagonal. The off-diagonal values range from near one to  $-0.26$ . The mean cross-correlation is  $0.3$ . This value is quite high, indicating that most of these post clusters are somehow positively related. We suspect this high average cross-correlation is at least partially due to external or systemic effects, which were not removed by the day-long bucket size. For instance, our analysis would be sensitive to patterns due to the Monday-Friday work week common in the West. Further, some of this correlation may be due to a high level of mutualistic interactions amongst the posts. Clearly, the ideas conveyed within the soc.women newsgroup often share similar contexts.

In our analysis, this overall high correlation does not particularly matter since we are concerned with the *relative* cross-correlation—that is, those that are the largest and those that are the smallest.

#### NEGATIVE CROSS-CORRELATIONS: COMPETITION VERSUS PREDATOR/PREY

We have primarily studied those pairs of quasi-species with relatively strong negative



**Figure 3** Volume of Activity for Two Quasi-Species

cross-correlations; to wit, those where  $\rho_{xy} \leq -0.2$ . Note that in all such cases (there are 42)  $p < .001$ , suggesting that with extremely high probability the correlations are not due to chance. Figure 3 plots two such interactions, both fairly characteristic of this population. [It] demonstrates a clear negative covariance between the two volumes of activity of the two post clusters. This negative covariance is both statistically significant and visually compelling. But what do these graphs signify and can it be interpreted within the rubric of ecological interactions?

At first glance the interactions appear to be of a predator/prey variety; they have a (+, -) relationship to them. However, competition might also produce similar interaction phenomena if the competitors are operating close to some limitation of environmental carrying capacity. In such instances, the relationship between population sizes will be a zero-sum game; when one goes up the other must come down. To be able to classify the interactions of Figure 3 we need to consider the qualitative details of these two interactions through direct study of the texts.

Recall that in the case of a predator/prey relationship, one population enjoys an increased growth rate at the *expense* of another population (e.g., one population feeds on the other). The presence of a relatively large population of predators will result in a diminished level of success for the prey (they

get eaten up). Conversely, the relative absence of prey will result in diminished success for the predator (they have nothing to eat).

Now consider the case of competition. In competition, two interacting populations inhibit each other in some way, reducing each other's level of success. This often occurs when the two populations rely on the same limited resource. Unlike the predator/prey relationship where the predator requires the prey for success, with competition the two populations would just as soon avoid each other all together.

This pressure towards avoidance is the source of much ecological diversity since it propels populations to explore new and therefore competition-free niches (Pianka, 1981). An ecological *niche*, for some particular species, is simply that collection of resources the species relies on. Interspecific niche overlap occurs when two or more species share one, some, or perhaps all of their resources. When those resources are scarce, interspecific competition will result. The *width* of a niche is simply a qualitative sense of the variety and number of resources a population makes use of.

### Competition and Niche Behavior

We have studied posts that make up the four quasi-species shown in Figure 3 in an attempt to qualitatively classify their interactions. The quasi-species (on the left side of Figure 3 are made up of posts within a single



thread. The subject line for these posts reads, "Men's Reproductive Rights." In general, these posts are concerned with the responsibilities and rights of men towards their unborn children. The quasi-species displayed with a dashed line in this part of the figure is centered on the use of contraceptives. It consists of a collection of posts wherein the authors debate who is most responsible, the woman or the man, when using contraception. The quasi-species with a solid line deals instead with the use of abortion and whether the father has any intrinsic rights in deciding whether or not to abort an unborn child.

[On the right side of] Figure 3, the two quasi-species are also from a single thread. The subject line here reads, "Unequal distribution of wealth?" This particular thread of discussion was rather large. In fact, there were a total of 365 posts to this thread, which our text analysis tools broke up into a number of quasi-species due to significant bifurcations of the topic. In other words, many parallel discussions occurred all within a single in-reply-to thread. The cluster of discussion shown with the solid line centered around a debate as to whether the US military was a "socialist collective." The quasi-species with the dashed line was a debate on the value of releasing the mentally ill from hospitals. Clearly, these two debates are quite dissimilar even though they span the same set of days and are posts to the same discussion thread.

The quasi-species [on the left in] Figure 3 are different but related discussions. Those [on the right] are different and not clearly related. Still, we believe that both of these sets of interactions demonstrate elements of competition. Within the texts, there is no evidence of predator memes; in fact, the memes seem entirely orthogonal to one another. However, in both examples the memes are competing for the same collection of human authors who must act as their agents if they are to propagate and succeed. This seems even more likely when we consider that all these posts are to the same newsgroup, which due to its narrow subject area supports only a limited supply of human posters. Moreover, each pair of interactions are confined to a single thread of discussion, which again has an even more

limited set of potential human authors since users of the NetNews system often zero-in on particular threads they find interesting and ignore others. After inspecting most of the interactions, which demonstrated strong negative correlations, we observed no examples of predator/prey interactions but many instances, which appeared to be examples of competition.

#### *Statistical Artifacts*

We computed the cross-correlation between 125 different clusters, arriving at 15,625 different correlations. It is possible, therefore, that the cross-correlations with large negative values exist simply by chance; they represent the tail of the distribution of correlations.

However, we believe that our qualitative analysis provides strong evidence that these negative correlations are *not* artifacts but are indeed due to an interaction phenomenon between the two quasi-species. The two pairs of quasi-species described in detail above demonstrate this point. The likelihood that two quasi-species would be brought together by mere chance *and* both be from the same thread (out of 324 threads within the corpus) seems vanishingly small.

#### *Competition*

We now will test our theory that these interactions are of a competitive nature. Again, recall that competition is often caused by populations existing within the same (narrow) ecological niche. What makes up an ecological niche for a meme within NetNews? We argue that the newsgroups themselves make up spatially distributed ecological niches. Since there is relatively little interaction between newsgroups (save the phenomena of cross-posting) we would expect these niches to behave something like island ecologies—they remain relatively isolated from each other. Within a single newsgroup (which is all we have studied so far) niches might be described by threads of discussions. As previously stated, we have found that individual posters to the system tend to become involved in particular in-reply-to threads that interest them. Thus, the memes within a

particular thread make use of a set of human resources, which is smaller than the entire set of potential human resources available to the newsgroup. These resources define the niche.

We theorize that cross-correlations that approach  $-1$  in our corpus are examples of competition, and competition will be more likely between populations that are posted to the same threads and thus have overlapping niches. The most direct way to test this theory is to see if negative cross-correlations between two quasi-species correlate with the degree to which they post to the same threads. For each of the  $125 \times 125$  pair-wise interactions we computed the number of threads each of the quasi-species pairs had in common and divided that by the total number of threads posted to by each quasi-species. For example, one quasi-species may contain posts that went to two different in-reply-to threads. Another quasi-species may have posts that span three different threads one of which is identical to a thread within the first group. So this pair of quasi-species would have posted to a total of four different groups one of which was shared. Their relative niche overlap would therefore be 0.25.

We calculated the correlation coefficient between the negative cross-correlations of Figure 2 and the percentage of thread overlap between these quasi-species pairs. We found this correlation to be  $-0.04$ . While this correlation is statistically significant ( $p < .001$ ), it is not very pronounced. The negative sign, though, does indicate that as the level of competition increases (a negative cross-correlation) the percent of overlap of their niche also increases (a larger positive shared thread percentage).

This small correlation coefficient may be due to a small signal/noise ratio. Since most pair-wise interactions result in small correlations, the relative number of large negative correlations is quite small. The number of interactions grows with the square of the number of quasi-species. We suspect that a simpler experiment, which grows linearly with the number of quasi-species, will have a better signal/noise ratio.

We have studied the correlations between the absolute number of in-reply-to threads a

quasi-species is posted to and the average degree to which the quasi-species finds itself correlated with other clusters. Our hypothesis is that the absolute number of threads a quasi-species is posted to will be related to the average degree of competition the quasi-species experiences in its interactions. Since the variety of resources used by an entity defines its niche, if a quasi-species is posted to a relatively small number of threads then it exists in a narrow ecological niche. Should there subsequently be any interspecific overlap of these narrow niches, scarcity will result in competitive encounters. We computed the correlation coefficient between the total number of threads within a quasi-species and its *average* cross-correlation value. The correlation coefficient here is 0.25. Thus, as the number of threads within a quasi-species increases (the set of available resources is widened) the average level of competition diminishes (the mean pair-wise cross-correlation also increases). This correlation is statistically significant ( $p < .001$ ) and rather pronounced.

We further computed the correlation coefficient when the absolute number of threads was normalized by the size of the quasi-species. We might expect that the number of threads employed by a quasi-species would grow with the number of posts within that quasi-species. In other words, as a quasi-species gets larger the number of threads increases too. This might affect the analysis above such that instead of measuring niche width we were simply measuring quasi-species size. Dividing out the size amounts to computing the *average* number of threads employed by a post for a given quasi-species. When this set of values was correlated with the mean cross-correlation, we arrived at a nearly identical coefficient as above and again clear statistical significance. Thus, quasi-species size is not a major factor in level of competition.

## CONCLUSIONS

We have described a set of text analysis tools, based primarily on Latent Semantic Indexing, which distill replicating memes from a corpus

of text. We have trained this analysis system on a corpus of posts to NetNews. This makes up a corporal ecology where the posts are organisms, NetNews is the environment, and human authors are a scarce resource. We argue that this represents an important bridging of text analysis and the Alife research program. Further, it amounts to a novel shift for Alife research—rather than synthesizing life-like agents, we are analyzing a pre-existing environment and discovering life-like behaviors.

In results reported here, we group together posts, which make use of similar sets of memes. These groups, clouds within a conceptual sequence-space, describe quasi-species. For each quasi-species, we compute its time-wise volume of activity by histogramming its daily post levels. We then study the pair-wise interaction between quasi-species by computing the cross-correlations between their time series. In our corpus, strong negative cross-correlations signify conditions of competition between the interacting populations where the quasi-species are competing for a limited set of human authors. Furthermore, quasi-species with relatively narrow ecological niches, those that make use of a small number of in-reply-to threads, are more likely to be in competition with other quasi-species. This behavior is analogous to what is found in natural ecologies (Pianka, 1981).

Why do these quasi-species compete? Qualitative analysis of the posts, such as those described in the previous section, shows that many competing quasi-species are posts sent to the same or similar threads. Competition is over the scarce authorship resources within these specific thread niches. Over time a particular thread of discussion may bifurcate into two or more internal themes which then proceed to compete for “air-time” within the thread.

## REFERENCES

- Berry, M. W. (1992). Large-scale sparse singular value computations. *International Journal of Supercomputer Applications* 6:13–49.
- Berry, M., O'Brien, T., Do, G., Krishna, V., & Varadhan, S. (1993). *SVDPACKC (Version 1.0) User's guide*. University of Tennessee Computer Science Department Technical Report, CS-93-194.
- Berry, M. W., & Fierro, R. D. (1995). *Low-rank orthogonal decompositions for information retrieval applications*. University of Tennessee Computer Science Department Technical Report, CS-95-284.
- Best, M. L. (1996). *An ecology of the Net: Message morphology and evolution in NetNews*. MIT Media Laboratory, Machine Understanding Technical Report, 96-001.
- Best, M. L. (1997). An ecology of text: Using text retrieval to study Alife on the net. *Journal of Artificial Life* 3:261–287.
- Chatfield, C. (1989). *The analysis of time series: An introduction*. London: Chapman and Hall.
- Dawkins, R. (1976). *The selfish gene*. New York: Oxford University Press.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41,6:391–407.
- Dumais, S. T. (1992). LSI meets TREC: A status report. In D. Harman (Ed.), *The First Text Retrieval Conference (TREC-1)*. NIST Special Publication 500-207.
- Dumais, S. T. (1993). Latent semantic indexing (LSI) and TREC-2. In D. Harman (Ed.), *The Second Text Retrieval Conference (TREC-2)*. NIST Special Publication 500-215.
- Eigen, M. (1992). *Steps towards life: A perspective on evolution*. Oxford: Oxford University Press.
- Eigen, M. J., McCaskill, J., & Schuster, P. (1988). Molecular quasi-species. *Journal of Physical Chemistry* 92,24:6881–6891.
- Foltz, P. W. (1990). Using Latent Semantic Indexing for information filtering. *Proceedings of the 5th Conference on Office Information Systems, ACM SIGOIS Bulletin*, 11,2–3: 40–47.
- Frakes, W. B., & Baeza-Yates, R. (Eds.). (1992). *Information retrieval: Data structures and algorithms*. Englewood Cliffs, NJ: Prentice Hall.
- Furnas, G. W., Deerwester, S., Dumais, S. T., Landauer, T. K., Harshman, R. A., Streeter, L. A., & Lochbaum, K. E. (1988). Information retrieval using a Singular Value Decomposition model of Latent Semantic Structure. *Proceedings of the 11th International Conference on Research and Development in Information Retrieval (SIGIR)*. New York: Association for Computing Machinery.

- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice Hall.
- Kantor, B., & Lapsley, P. (1986). Network news transfer protocol: A proposed standard for the stream-based transmission of news. Internet RFC-977.
- Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics* 1:1–18.
- May, R. M. (Ed.). (1981). *Theoretical ecology principles and applications*. Oxford: Blackwell Scientific Publications.
- Mettler, L. E., Gregg, T. G., & Schaffer, H. E. (1988). *Population genetics and evolution* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Pianka, E. R. (1981). Competition and niche theory. In R.M. May (Ed.), *Theoretical ecology principles and application* (pp. 167–196). Oxford: Blackwell Scientific Publications.
- Pielou, E. C. (1969). *An introduction to mathematical ecology*. New York: Wiley-Interscience.
- Pocklington, R. (1996). *Population genetics and cultural history*, Masters Thesis. Burnaby: Simon Fraser University.
- Pocklington, R., & Best, M. L. (1997). Cultural evolution and units of selection in replicating text. *Journal of Theoretical Biology* 188:79–87.
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24,5:513–523.

# 3.7

## IDENTIFYING THE UNKNOWN COMMUNICATOR IN PAINTING, LITERATURE AND MUSIC

WILLIAM J. PAISLEY\*

**T**he task of identifying the author of an anonymous work has long challenged students of communication. The problem is usually posed in one of three ways:

(1) A work is attributed to a communicator well known for other works (the generic “communicator” here denotes painter, writer, composer), but it may be an imitation or forgery (e.g., the “Corelli” violin sonatas by Fritz Kreisler)

(2) A truly anonymous work is attributed by default to a well-known communicator whose own works are similar, but it may be the work of a disciple or lesser-known colleague (e.g., the “Letter to the Hebrews,” long attributed to Paul)

(3) A work is attributed variously to each of two or more well-known communicators (e.g., the perennial Shakespeare-Bacon-Marlowe-Oxford wrangle)

Research directed to this problem has two goals. The practical goal is a correct attribution of the anonymous or disputed work. The theoretical goal is a better understanding of one phase of the communication process, the encoding of messages.

When authorship of a work is disputed, we may assume that historical evidence is inadequate. Therefore, clues must be sought in the text itself, usually in the style of the work. “Style,” however, is a concept often embracing the ineffable qualities of a communicator’s output. To focus on objective characteristics of the text, a concept such as “encoding habits” should be substituted for “style.” Then the unique character of a work may be defined in terms of successive decisions made by the communicator as he chooses from his repertory of symbols (notes, words, brush strokes, etc.).

In the last quarter of the 19th century, when the connoisseur of art finally rejected

\*From Paisley, W. J. (1964). Identifying the unknown communicator in painting, literature and music. *Journal of Communication* 14, 4:219–237.

dubious historical evidence and turned his attention to the encoding habits of painters as preserved in their works, he achieved a great refinement in the technique of connoisseurship. Recent successful efforts to identify the authors of anonymous literary works reveal a refinement of method in that field also. Although art connoisseurship has remained a quantitative science while literary detection has become exhaustively quantitative, nevertheless there is surprising consensus in both fields concerning those encoding habits which clearly distinguish a communicator from all other communicators with superficially similar output. This consensus is interesting insofar as it defies common sense and favors *minor* encoding habits which are inconspicuous in the work and do not carry the burden of meaning. Since the communality of the two streams of research has perhaps never been discussed, the first half of this report will summarize their shared assumptions and procedures. The second half reports the results of an extension of these procedures to the study of musical encoding habits.

#### THE CONNOISSEUR AS CONTENT ANALYST

In 1889 the best-known connoisseur of this century, Bernard Berenson, visited Rome and discovered a new way of looking at paintings. He has left a memoir of the experience:

A generation ago, when a beginner, I enjoyed the privilege of being guided through the Borghese Gallery by a famous connoisseur. Before the *Pieta* now ascribed to Ortolano I fell into raptures over the pathos of the design. My mentor . . . cut me short with, "Yes, yes, but please observe the little pebbles in the foreground. They are highly characteristic of the artist." "Observe the little pebbles" has become among *my* intimates a phrase for all the detailed, at times almost ludicrously minute, comparisons upon which so large a part of activities like mine are spent. (Kiel, 1962: 145–146)

Berenson's mentor was Giovanni Morelli, whose "scientific connoisseurship" provoked

international controversy in art circles. Whereas his colleagues were content to accept the testimony of Vasari and other early historians of art, Morelli contended that paintings sufficiently identified themselves, each work signed by its creator in dozens of little details, which no two painters executed alike. Traditional criteria based on the overall style of a work were entirely misleading, he claimed, since the student in any Renaissance studio soon learned the superficial marks of his master's style.

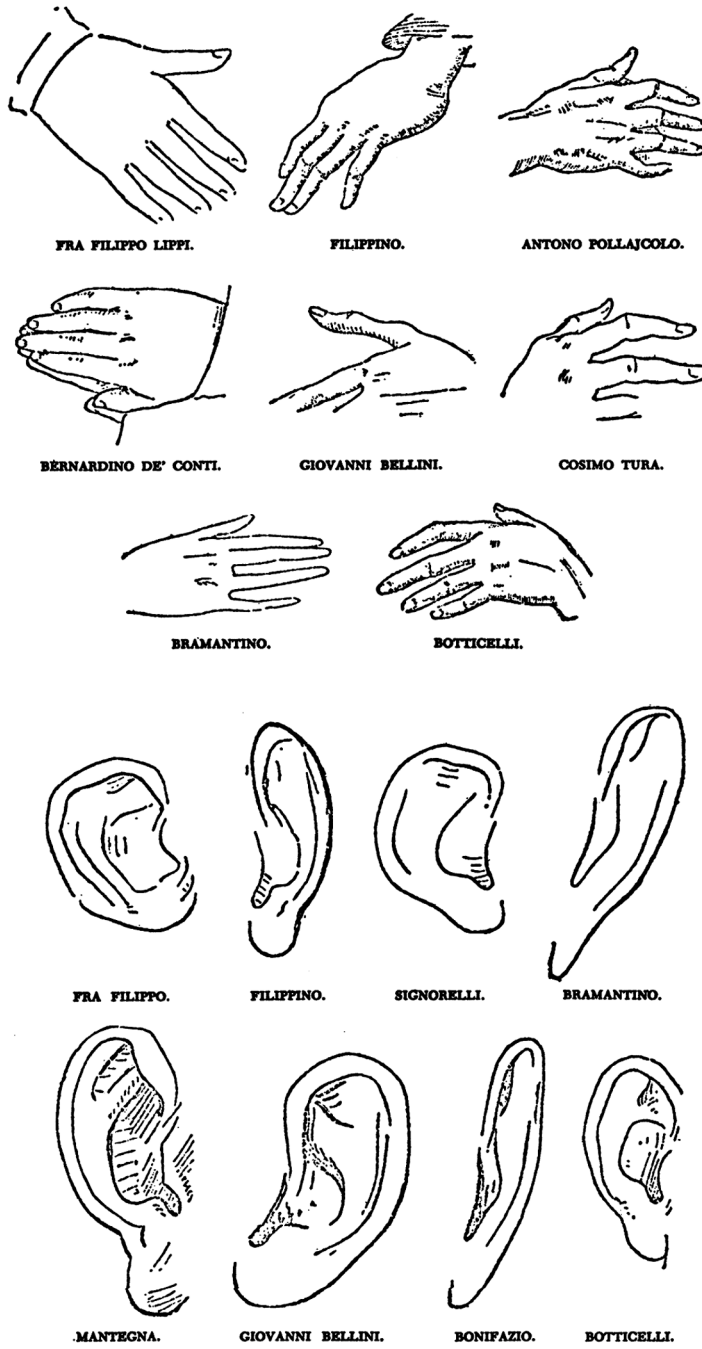
Morelli was at heart a taxonomist, as Fig. 1 and the following excerpts illustrate:

Look at the Raphaellesque type of ear in the children; see how round and fleshy it is; how it unites naturally with the cheek and does not appear to be merely stuck on, as in the works of so many other masters; observe the hand of the Madonna with the broad metacarpus and somewhat stiff fingers, the nails extending to the tips only. (Morelli, 1900:37)

Among Sandro Botticelli's characteristic forms I will mention the hand, with bony fingers—not beautiful, but always full of life; the nails, which, as you perceive in the thumb here, are square with black outlines. (Morelli, 1900:35)

Such attention to minor detail brought derision from Morelli's colleagues. They did not understand his distinction between "appreciating" a work and studying a painter's encoding habits. Morelli was called "the connoisseur of fingernails." Yet the success of his method could not be ignored [he exposed scores of mislabeled Renaissance works in Italian and German galleries, 46 in the Dresden Gallery alone (Wind, 1964:29)], and connoisseurship gradually committed its future to content analysis. Bernard Berenson has summarized the assumptions of the new connoisseurship:

Obviously, what distinguishes one artist from another are the characteristics he does not share with others. If, therefore, we isolate the precise characteristics distinguishing each artist, they must furnish a perfect test of the fitness or unfitness of the attribution of a given work to a given master. (Berenson, 1902:123–124)



**Figure 1** Hands and Ears Sketched by Giovanni Morelli to Illustrate Idiosyncrasy in the Execution of Minor Details by Renaissance Painters

SOURCE: Morelli (1900:77-78).

## THE LITERARY DETECTIVE AS CONTENT ANALYST

The venerable “who was Shakespeare?” controversy provides a mirror in which the changing character of literary sleuthing can be traced. If it happened that connoisseurs of art turned to content analysis when they found historical evidence inadequate, then certainly theft colleagues in literature turned to content analysis as the only alternative to a conspiratorial view of history. That is, the modern champions of Bacon, Marlowe, the Oxford group and other contenders for the crown have admitted that history indeed appears to affirm that the playwright and the actor Shakespeare are one man, *but* history is deceitful and by clever subterfuge the plays of Bacon (or the Oxford group, or the exiled Marlowe) were performed under the name of the dull-witted and mercenary Stratfordian. Courtly discretion (they assert) led Bacon and Oxford to avoid association with the vulgar stage, while Marlowe thought it necessary to pretend that he was dead. When scholars hatch conspiracies, then content analysis of the plays themselves is clearly the only recourse.

Unfortunately, this Elizabethan imbroglia may not be solved, even by content analysis. Of all the contenders, only the least likely, Marlowe, left plays, which permit the necessary comparison of encoding habits. Mendenhall (1887) thought it legitimate to use Bacon’s civil and political essays as a corpus for comparison, but no modern scholar would insist that a communicator’s encoding habits must remain constant between works as topically and structurally remote as the essays and the plays. If a striking consistency had been found between essays and plays in Mendenhall’s analysis of word-length frequencies, then Bacon’s claim would have been strengthened. The inconsistencies actually found are evidence of nothing.

If content analysis is not likely soon to disclose who Shakespeare was, nevertheless other applications of the technique in authorship identification have been successful.

Moreover, over time the focus of these efforts has shifted steadily from major to minor encoding habits. Three representative efforts will be discussed in chronological order:

### (1) Yule’s study of Gerson, à Kempis, Macaulay and others.

George Udny Yule contributed a great deal more to the study of literary vocabulary than can be acknowledged here, but at least his study of the *Imitatio Christi* must be described. Because the *Imitatio* had been ascribed both to Gerson and to Thomas à Kempis, Yule first compiled a large sample of undisputed works by each man. Then in the *Imitatio*, he chose words, chiefly nouns, which seemed to be favored by the unknown author. Finally, he tabulated frequencies of occurrence for these words not only in the *Imitatio* but also in the two comparison samples. When the *Imitatio* distributions were arrayed beside those of Gerson and Thomas à Kempis, it could not be doubted that à Kempis was the author.

Yule’s perseverance in this most tedious research was remarkable, but his concentration on *major* encoding habits may have increased his labor while reducing the sensitivity of his measure. Thus use of the noun “prayer” involves a major encoding (in this context major encoding habits are those that carry the burden of meaning), while use of the article “the” involves a minor encoding habit. Unlike the ubiquitous “the,” “prayer” is a relatively rare word; large samples of text must be scanned to obtain minimum stable frequencies. Moreover, because “prayer” is topic-constrained (found in texts with a religious topic), texts on different topics *by the same author* may yield anomalous frequencies of “prayer.”

In his comparison of three essays (“Milton,” “John Hampden,” “Frederick the Great”), Yule (1944:122) shows—perhaps inadvertently—the effect of topic differences on major encoding habits. The most common nouns in each essay are (in descending order):



<i>Milton</i>	<i>Hampton</i>	<i>Frederick</i>
Man	King	King
Poet	parliament	man
Character	man	army
Poetry	house	war
Mind	time	time
Time	commons	year
Work	people	prince
People	member	part
Liberty	year	power
Power	party	day

Although all three essays are biographies of public figures by the highly idiosyncratic Macaulay, these lists lack authorship communality. Sixteen words are found in one list only; four words in two lists; only two words in all three lists. Of all literary encoding habits, the use of nouns is perhaps most constrained by topic differences.

In the course of this research Yule himself considered, and decided against, the study of minor encoding habits:

Would it be of service to include other words (in addition to nouns, adjectives, and verbs)? If so, should it be *all* other words or should there be some specified exceptions, such as say the definite article in English, or auxiliary verbs? My impression is that the inclusion of all words without exception would be a mistake; that the inclusion of *a* and *the* and *is* and *the* like, each with a very large number of occurrences in any author, would merely tend to obscure differences, and it would be best to limit data to what are in some sense “significant words.” (Yule, 1944:280)

Two recent studies indicate that Yule’s bias against “insignificant words” was unjustified.

## (2) Ellegard: the “Junius” letters.

Ellegard (1962) incorporated many of Yule’s procedures in his investigation of the authorship of the “Junius” letters (published over that pseudonym between 1769 and 1772 in the *London Public Advertiser*). Alert, however, to the noun-counting trap, which cost Yule so much labor, Ellegard limited his own selection of words to abstract nouns, adjectives, adverbs and prepositional constructions.

Without explaining his strategy in these terms, Ellegard nonetheless rejected Yule’s focus on major encoding habits.

There are five steps in Ellegard’s procedure:

- (i) The disputed text is scanned for words that are conspicuously frequent. These are “plus words” for that text
- (ii) A large body of contemporary writing is scanned for words which are conspicuously absent in the disputed text. These are “minus words” for that text
- (iii) When lists of “plus words” and “minus words” have been compiled, all texts are scanned again and exact frequencies of occurrence are recorded. These counts may show that some words are not as “plus” or as “minus” as had been supposed, and the list may have to be revised
- (iv) All likely authors of the disputed text are sampled. Their texts are counted for frequencies of “plus” and “minus” words
- (v) That author whose word-profile most nearly resembles that of the pseudonymous author is the probable choice, provided that other candidates’ profiles are significantly dissimilar (i.e., their counts regularly fall outside the confidence limits established for the many samples of disputed text)

This painstaking procedure permitted Ellegard to conclude that Sir Philip Francis, historically regarded as the author of the “Junius” letters, is a best choice on statistical grounds also.

## (3) Mosteller and Wallace: the Federalist papers.

In 1963, about 80 years after Morelli observed the stable idiosyncrasy or minor encoding habits in painting, Mosteller and Wallace reported the same phenomenon in literature. Morelli ignored major encoding habits because he assumed that any competent forger could fool him at that level. In their study of twelve disputed Federalist papers, Mosteller and Wallace chose to ignore major encoding habits for two reasons: (i) the two authors in question, James Madison and Alexander Hamilton, conformed their writings to the intricate and formal rhetoric of the time—hence superficial differences cannot be

found; (ii) the disputed Federalist papers discuss a wide range of topics—hence authorship differences and topic differences may be confounded in the case of semantically significant words.

Mosteller and Wallace define their focus in these terms:

As we implied in discussing the word war, the words we want to use are noncontextual ones, words whose rate of use is nearly invariant under change of topic. For this reason, the little filler words, called function words, are especially attractive for discrimination purposes. (1963:280)

Within the class of function word (as distinguished from content words, cf. Fries, 1952) there is great variation in frequency of occurrence. After they had tested each function word for low variance within an author's works and high variance between the works of the two authors, Mosteller and Wallace divided their list of words into sets according to frequency. They found that the highest-frequency set (consisting of the words *to*, *there*, *on*, *of*, *by*, *an* and *also*) discriminated as powerfully as the three lower-frequency sets taken together. Remarkably, the preposition *upon* proved to be a reliable discriminator by itself, since Hamilton used it five times more often than Madison.

Calculating odds from these discriminations, Mosteller and Wallace were able to attribute all the disputed papers to Madison. The least distinctive paper gives Madison 80 to 1 odds; the next weakest gives him 800 to 1 odds; thereafter the odds become astronomical.

In summary, Mosteller and Wallace relied on the "insignificant" words which Yule thought too ubiquitous to distinguish between authors. As a result they achieved discrimination (when the two sets of known papers are compared) with median odds of *3 million to 1*.

#### A WORKING DEFINITION OF "MINOR ENCODING HABITS"

The distinction between major and minor encoding habits has been viewed from two perspectives, which may now be combined.

Morelli and Berenson introduced four defining criteria:

- (1) The detail to be studied should not be prominent; else imitators will appropriate it
- (2) It should be executed mechanically (i.e., with little feedback for self-criticism); else the communicator may consciously vary it for effect
- (3) Its use should not be dictated wholly by convention (e.g., the halo in Renaissance paintings)
- (4) It should not be so rare that examples cannot be found in each disputed work

To this list, Mosteller and Wallace would add:

- (5) The detail should remain constant in frequency whatever the topic of the work; else topic difference will confound authorship differences

But their quantitative procedure requires a restatement of other criteria:

- (2a) Use of the detail should exhibit low variance within a communicator's works
- (3a) Use of the detail should exhibit low variance within a communicator's works
- (4a) Frequency of occurrence should be high relative to the sampling error

These criteria have helped to identify unknown communicators in painting and literature. It would be desirable to report their relevance to the study of authorship differences in music, the third leg of the triangle of "artistic" communication, but that research has yet to be done. The findings reported below concern only one of many aspects of musical communication, which invite investigation.

#### IDENTIFYING THE COMMUNICATOR IN MUSIC

In the encoding process called musical composition, certain variables can assume many states and therefore require successive

choices on the part of the communicator. Among these are: (1) tempo, (2) dynamics, (3) harmony, (4) instrumentation, (5) pitch. Each variable implies a set of encoding habits, major or minor, and each may prove to discriminate reliably between composers. However, it is within the scope of the present study to consider only the variable of pitch, of note-to-note pitch transitions in the themes of selected composers. This variable has been chosen for two reasons: (1) changes in pitch are easily coded for processing by computer; (2) some research involving tonal transitions has already been reported.

Previous studies of tonal transitions have been guided either by information theory or by the probability model of stochastic processes. The measures employed in the information-theory studies are those of uncertainty and constraint, each composer or sample of works described by two or three parameters only (cf. Youngblood, 1958). The stochastic process studies (cf. Brooks et al., 1957) yields matrices of transitional probabilities over 1, 2, 3 . . . n steps, a rich mine of information about the composer's encoding habits.

Unfortunately, no study reports testable data on differences between composers. Lacking replication within the works of each composer, variance terms necessary for testing the inter composer difference cannot be computed. This is not to be construed as a defect in method, of course, since the investigators were not seeking to test such differences. They sought single best estimates for each composer (or for each sample of melodies) either of uncertainty or of the probability of a given transition, and these estimates they obtained.

Therefore, this study focuses on a variable that, although not entirely unresearched, varies in yet-undetermined patterns between and within composers' works.

### Problem

Bernard Berenson said that connoisseurship "proceeds, as scientific research always does, by the isolation of the characteristics of the known and the confrontation with the unknown" (Berenson, 1902:124). This study seeks to isolate, in the themes of five

composers (Bach, Haydn, Mozart, Beethoven, Brahms), those minor encoding habits, which identify each man. Then, given the characteristics of the known, four "unknown" samples are tested to determine whether they may be accepted or rejected as the work of any of the five composers. Since the four test samples are actually of known authorship (Handel, Mozart, Beethoven, Mendelssohn), the validity of the discrimination technique may be assessed. Yet the rigor of the test is preserved by setting aside the "unknown" samples and leaving them unanalyzed until the discriminating characteristics of the five composers have been established.

**Procedure.** The collection of hundreds of compositions, the transcription of sections from them and the transposition of these sections to a common tonality for comparison are a task, which only a team of well-financed investigators could undertake. Fortunately, for the fate of the present modest study, a source was found in which this work has already been done. Barlow and Morgenstern (1948) have indexed about 10,000 themes from the works of dozens of composers after transposing each theme from its original tonality to the keys of C major and C minor. Since each theme is represented in the index as a sequence of letters (e.g., E F B C F# F#—the first theme of Beethoven's First Symphony), it is a relatively simple task to keypunch and then code the samples of themes for computer processing.

The major composers named above were chosen for this study because Barlow and Morgenstern have indexed particularly large reactions of their themes. Altogether, 2,240 themes were sampled systematically from this source. Because the index consists only of those opening notes of each theme, which clearly identify it, only the first six notes from each theme (the minimum entry in the index) could be keypunched. This is a waste of data in those frequent instances in which it takes as many as 12 notes to identify a theme, but sampling consistency is the more important consideration. The entire data of this study therefore consist of 13,340 notes, divided as follows: Bach, 1,920; Handel, 960; Haydn,

1,920; Mozart, 2,880; Beethoven, 2,880; Mendelssohn, 960; Brahms, 1,920.

The samples from Handel and Mendelssohn (160 themes each) were set aside as “unknowns.” Random samples of 160 themes from Mozart and Beethoven were also set aside for later testing. The remaining 320 themes from each of the five comparison composers were randomly divided into half-samples of 16 themes in order to permit the estimation of variance within the works of each composer.

A first step in the computer processing (performed on the Stanford University 70901) was the recoding of letters as numbers. For this purpose a twelve-tone scale is a more efficient model than a diatonic scale, and the letters received the following values: C, 1; C# and Db, 2; D, 3; D# and Eb, 4; E, 5; F, 6; F# and Gb, 7; G, 8; G# and Ab, 9; A, 10; A# and Bb, 11; B, 12. In a later analysis the twelve categories were reduced to five: the tonic (1), the third (5), the fifth (8), all other diatonic tones (3,6,10,12) and all chromatic tones (2,4,1,9,11). This collapsing yields more equal proportions in each category by taking account of the do—mi—sol trimodality of this music.

### Results: First Analysis

It was decided to look first at the simplest transitions and subsequently, if necessary, to

study also the more complex ones. Greatest idiosyncrasy is found in greatest complexity, of course, since no two composers have ever assembled even a hundred notes in quite the same sequence, but idiosyncrasy found in complexity is not generalizable to other works by the same man—as other composers do not repeat the sequence, neither does he.

If two-note transitions are classified in terms of the identities of the first and second notes, 144 ( $12 \times 12$ ) categories result. As the number of original categories is reduced by collapsing (say from 12 to 5), then the number of joint classifications diminishes in proportion to the square of the number of categories, to 25. There is great economy in reducing all pitch encoding decisions to 25, but even greater simplicity may be achieved by sacrificing pitch identities and considering only the size of the interval separating the first and second notes. Since two notes can be separated by no more than six semitones (e.g., the distance from C to fl in either direction), the number of categories needed is only seven (from 0 to 6 semitones). Thus, the 800 2-note transitions in each 160-theme sample may be coded just seven ways.

Table 1 shows the tabulation of “jumps” (so called to avoid the term “interval,” which suggests vertical harmony rather than horizontal motion) for the ten samples of the five composers.

**Table 1** Frequency With Which The Five Composers “Jump” From 0 to 6 Semi-Tones in Each of Their 160-Theme Samples

Number of Semi-Tones “Jumped” in Each Transition	<i>Bach</i>		<i>Haydn</i>		<i>Mozart</i>		<i>Beethoven</i>		<i>Brahms</i>	
	1	2	1	2	1	2	1	2	1	2
0	60	86	141	150	163	164	147	136	76	100
1	200	199	175	177	172	168	159	182	193	211
2	264	269	229	196	203	191	216	222	235	225
3	90	80	98	106	100	99	96	99	113	112
4	72	62	75	69	71	76	81	70	73	52
5	113	103	82	100	86	101	98	89	104	93
6	2	2	1	3	6	2	4	3	7	8

NOTE: Number of 2-note transitions in each sample is 800.

Since pitch identities have been lost, category 0 represents such transitions as C-C, C#-C#, D-D, etc. Category 2, the modal category in all ten samples, represents such whole-tone (two semitone) transitions as C-D, C#-D#, D-E, D#-F, etc. Three facts are immediately apparent in Table 1: (1) composers agree roughly in the frequency with which they use each “jump,” (2) there is error variance between the two samples of each composer, (3) yet the two samples of each composer tend to vary less around their own mean than around the ten-sample mean. This last fact is a test of the discriminatory power of these simple categories in both the known and the “unknown” samples.

The appropriate statistic is the chi-square goodness-of-fit test.<sup>1</sup> Theoretical frequencies are defined as the means for each composer’s two samples of the six categories (six rather than seven, with 5 and 6 aggregated, because the theoretical frequencies for category 6 by itself would be insufficient for chi-square). Table 2 reports chi-squares obtained for the original samples and also for the “unknown” samples. Values entered on the major diagonal are in effect error terms—chi-squares of the extent to which each composer deviates from his own mean. With 5 degrees of freedom, chi-squares of 11.1, 15.1 and 20.5 have probabilities of .05, .01, and .001, respectively. Thus in the matrix of known samples Bach and Brahms each reject, and are rejected by, all composers

except himself. This pattern persists in the group of “unknown” samples, the probability being extremely small that either Bach or Brahms could have written any of the four.

Unfortunately, Haydn, Mozart and Beethoven cannot be distinguished on the basis of “jumps.” Discrimination of the three classical composers is not only weak; it is anomalous in that Haydn accepts Beethoven’s known samples with less error than his own. Therefore, the “jumps” analysis separates Bath and Brahms from the group and establishes that neither could have written the “unknown” samples, but the three contemporaries cannot be found to differ in this encoding habit.

### Second Analysis

The next simplest classification of two-note transitions has already been described: 25 joint classes based on the tonic, the third, the fifth, all other diatonic tones and all chromatic tones. Accordingly, the ten known samples were processed again on the computer to provide the frequency distributions reported in Table 3. The patterns observed in Table 1 reappear in this table: error variance within a composer’s works, but less variance around each two-sample mean than around the ten-sample mean.

The tables have other affinities. In Table 1 it was seen that Mozart was the composer most likely to repeat a note (i.e., make a

**Table 2** Goodness-of-Fit Chi-Squares Obtained When the Mean “Jump” Frequencies for Each Composer Are Taken as Expected Frequencies and the 14 Sets of Sample Means Are Taken as Observed Frequencies

<i>Source of Expected Frequencies</i>	<i>Known Samples<sup>a</sup></i>					<i>Unknown Samples<sup>b</sup></i>			
	<i>Bach</i>	<i>Haydn</i>	<i>Mozart</i>	<i>Beethoven</i>	<i>Brahms</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
Bach	3.2	94.6	139.7	83.2	21.8	78.1	164.2	49.9	89.5
Haydn	61.3	2.8	4.5	2.7	35.5	6.2	8.5	10.2	4.2
Mozart	85.9	6.5	0.6	7.2	53.7	11.4	4.2	21.2	9.4
Beethoven	54.8	3.7	6.3	1.7	35.3	3.3	13.6	6.6	3.6
Brahms	18.4	48.9	79.3	45.2	4.1	48.2	90.0	27.5	51.0

a. Each entry in these cells is derived from the mean for the two samples of each composer.

b. The four “unknown” composers are Handel, Mozart, Beethoven and Mendelssohn.

0-semitone “jump”). In Table 3 it is Mozart who most often uses the repetitions transitions (tonic—tonic, third—third, fifth—fifth). In Table 1 Bach and Brahms were highest in 1-semitone “jumps,” which in ten out of twelve cases involve chromatic tones (E-F and B-C being the only 1-semitone

intervals in the diatonic scale of C major). Therefore, it is not surprising that Bach and Brahms lead the group in all chromatic transitions in Table 3. Other examples may be found in which the tables mutually support the patterns of encoding habits found in each.

**Table 3** Frequency of 25 Types of Two-Note Transitions in Each of the Ten 160-Theme Samples From the Five Composers

<i>From-To</i>	<i>Bach</i>		<i>Haydn</i>		<i>Mozart</i>		<i>Beethoven</i>		<i>Brahms</i>	
	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>
Tonic to tonic	9	23	46	63	64	74	34	33	19	17
Tonic to third	23	19	24	28	31	41	22	22	11	12
Tonic to fifth	47	37	23	31	25	35	22	24	25	19
Tonic to other diatonic	88	101	80	93	69	70	97	81	62	79
Tonic to all chromatic	24	20	11	8	12	14	10	16	32	21
Third to tonic	9	14	26	21	15	21	25	17	10	17
Third to third	6	11	20	17	29	18	23	42	18	26
Third to fifth	15	15	11	28	22	26	17	17	25	19
Third to other diatonic	50	44	80	53	69	63	52	54	63	46
Fifth to tonic	52	52	36	53	45	40	44	40	40	30
Fifth to third	14	23	30	31	32	35	27	18	23	21
Fifth to fifth	31	28	48	55	51	59	58	38	16	36
Fifth to other diatonic	65	59	50	56	50	52	54	57	57	61
Fifth to all chromatic	33	34	9	13	19	18	14	27	41	41
Other diatonic to tonic	52	63	62	69	49	59	63	65	56	48
Other diatonic to third	36	29	54	44	43	44	40	45	43	42
Other diatonic to fifth	51	51	48	32	42	33	49	39	41	35
Other diatonic to other diatonic	37	48	77	64	76	35	86	75	47	67
Other diatonic to all chromatic	56	38	19	9	13	14	17	19	44	36
All chromatic to fifth	25	28	8	12	21	19	6	23	35	41
All chromatic to other diatonic	51	42	22	12	16	17	20	22	53	33
All chromatic to all chromatic	8	15	9	5	3	7	10	9	23	25
<sup>a</sup> All chromatic to tonic and third	21	6	7	3	4	6	10	17	16	27
<sup>a</sup> Third to all chromatic										

a. Aggregated to provide sufficient expected frequencies for chi-square analysis.

**Table 4** Goodness-of-Fit Chi-Squares Obtained When the Mean Transitional Frequencies for Each Composer Are Taken as Expected Frequencies and the 14 Sets of Sample Means Are Taken as Observed Frequencies

Source of Expected Frequencies	Known Samples <sup>a</sup>					Unknown Samples <sup>b</sup>			
	Bach	Haydn	Mozart	Beethoven	Brahms	1	2	3	4
Bach	16.2	295.3	376.7	229.4	110.7	157.9	414.6	181.9	148.6
Haydn	355.5	20.0	51.9	82.5	468.2	116.3	55.5	78.8	236.5
Mozart	324.3	51.2	14.3	100.3	386.0	127.2	24.2	65.9	180.6
Beethoven	201.3	65.6	103.0	17.3	214.1	91.1	114.6	36.5	102.7
Brahms	94.9	278.7	351.1	155.0	17.4	195.4	343.1	122.8	143.0

a. Each entry in these cells is derived from the mean for the two samples of each composer.

b. The four "unknown" composers are Handel, Mozart, Beethoven and Mendelssohn.

Table 4 reports chi-squares obtained when the goodness-of-fit test is applied to means computed from the data of Table 3.

Since there are 23 means for each composer (3 of the 25 classes having been aggregated to provide sufficient frequencies), the resulting chi-squares must be evaluated with 22 degrees of freedom. The .05, .01 and .001 probability values of chi-square are 33.9, 403 and 483, respectively.

Unlike Table 2, Table 4 shows sharp discrimination in the matrix of known samples. The chi-squares on the major diagonal, again measures of inter-sample error, give each composer at least 50–50 odds that he actually wrote his own works, while the remaining chi-squares in this matrix give odds exceeding 1,000 to 1 that each composer did not write the known samples of the other four composers. As before, Bach and Brahms are separated from the other composers by extremely high chi-squares, but even the Haydn-Beethoven distinction is sharply drawn. The weakest discrimination, between Haydn and Mozart, easily meets a 1,000–1 criterion of rejection.

With the assurance that this set of encoding habits actually does discriminate, the four "unknown" samples may be tested again. Chi-squares resulting from this phase of the analysis show immediately that *none* of the five composers could have written samples 1 and 4. Nor, if a .001 criterion is

established, could Bach, Haydn or Brahms have written *any* of the four. Nor could Beethoven have written sample 2, nor Mozart sample 3. But sample 2 could easily represent a chance deviation from Mozart's mean frequencies, while sample 3 suggests a more outlying but quite possible deviation from Beethoven's mean frequencies. Foreknowledge of the authorship of samples 2 and 3 may dispose the investigator to attribute them to Mozart and Beethoven, but it seems that the chi-squares also speak for themselves.

### Summary of the Analyses

It was decided to begin with the simplest classifications of encoding habits and to seek the least complex behavior, which would prove to be reliably idiosyncratic. Whereas it had been expected that analyses would be required of 3-note, 4-note and perhaps even higher-order transitions, a classification of 2-note transitions into only 25 categories satisfied a stringent discrimination criterion and led to the proper disposition of the "unknown" samples.

### DISCUSSION

As a contribution to communication research, this study scarcely ranks with the precedent

studies reviewed above. In the first place, it lacks their research problem—an unknown communicator to identify. Secondly, modest amounts of data are involved. Thirdly, computer processing eliminates most of the tallying and testing which attained awesome proportions in Yule's work.

Yet, acknowledging these differences, this study seems to close the triangle of "artistic" communication by establishing that composers too have their minor encoding habits—analogs of the writer's prepositions and the painter's fingernails. Although only the variable of pitch was studied, it seems safe to infer that composers also differ at this microanalytic level in their use of rhythms, harmonies, etc.

Indeed, converging evidence now suggests that all human communicative behavior exhibits two types of idiosyncrasy deserving study. The first is the obvious idiosyncrasy of complex constructions. That is, no two men could independently encode *Paradise Lost* or even three stanzas from it. Even the single sentence "Him the Almighty Power/Hurled headlong flaming from the eternal sky/With hideous ruin and combustion down/To bottomless perdition, there to dwell/In adamant chains and penal fire/Who durst defy the Omnipotent to arms" would keep the fabled chimpanzees at their typewriters for centuries. The second type of idiosyncrasy is that of minor encoding habits, which lie at an opposite pole from complex constructions on the continua of deliberation and self-consciousness. For instance, whether he was aware of his proclivity or not, Fra Filippo liked to render an earlobe as circular while Bonifazio preferred an elongated ellipse (see Fig. 1). Mozart liked to repeat a tone in consecutive notes while Bach preferred to move up or down a semitone. If asked, could these communicators state why they had chosen certain patterns and not others?

It is tempting to link such behaviors to unconscious determinants and thus escape responsibility for explaining them. But the behaviors under consideration here (e.g., the use of prepositions) are as devoid of affect as human activity can be, and motivation for unconscious determination is therefore

lacking. It would be absurd to argue that Hamilton felt impelled to use upon or that Madison felt impelled to censor his use of it.

A more satisfactory perspective is that of learning theory. Having at some time, somehow, been reinforced for using upon, Hamilton continued to use it more frequently than did his contemporary Madison, who may or may not have been reinforced for avoiding it. We cannot yet infer what the relevant reinforcements might have been, but we may certainly infer that selective reinforcement was involved and that behavior was "shaped" to this end.

Common sense supports the assertion that trivial details are subject to random variation while significant details are frozen in the mold of the communicator's intention. Evidence now suggests, however, that no detail is so trivial that it does not vary systematically within and between communicators' works. Many studies, most recently this one, have asked *how?* Perhaps the next will ask *why?*

## NOTE

1. Chi-square provides an estimate of the probability that a given set of departures from theoretical frequencies could have occurred by chance. Unlike tests based on the standard error of the mean (*t* test, analysis of variance), chi-square requires no assumptions of normality and variance homogeneity and is therefore applicable in situations (such as this) in which each mean is computed from only two values and in which information about the sampling distribution is unavailable.

## REFERENCES

### Painting

- Berenson, B. (1902). *The study and criticism of Italian art*. London: G. Bell and Sons.
- Kiel, H. (Ed.). (1962). *The Bernard Berenson treasury*. New York: Simon and Schuster.
- Morelli, G. (1900). *Italian painters* (Vol. 1, C. J. Ffoulkes, Trans.). London: John Murray.
- Wind, E. (1964). Critique of connoisseurship. *Art News* 63:26–29, 52–55.



### Literature

- Ellegard, A. (1962). *A statistical method for determining authorship*. Goteborg, Sweden: Elanders Boktryckeri Aktiebolag.
- Fries, C. (1952). *The structure of English*. Ann Arbor: University of Michigan Press.
- Mendenhall, T. C. (1887). The characteristic curve of composition. *Science* 9, 214:237–246.
- Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem. *Journal of the American Statistical Association* 58: 275–309.

- Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge, England: Cambridge University Press.

### Music

- Barlow, H., & Morgenstern, S. (1948). *A dictionary of musical themes*. New York: Crown.
- Brooks, F. P., Jr., et al. (1957). *An experiment in musical composition*. I.R.E. Trans. Elec. Computers, EC-0:175.
- Youngblood, J. E. (1958). Style as information. *Journal of Music Theory* 2:24.

# 3.8

## WHEELS OF TIME AND THE INTERDEPENDENCE OF VALUE CHANGE IN AMERICA

J. ZVI NAMENWIRTH\*

... It is my contention that the history of value change is neither progressive nor regressive, but basically cyclical. I shall therefore try to demonstrate the plausibility of this assertion; relate cycles of change in a variety of values, thereby delineating the underlying structure of the cyclical findings, or "The Wheel of Time"; attempt to interpret the meaning of the wheel; and then conclude with a speculation about its possible causes.

### VALUES: CONCEPT AND ASSESSMENT

The definition and assessment of value change determine to some extent the findings, and an explication is therefore in order. For this exposition, the distinction between goods and values is basic. Goods are the available resources of a society at any one time. These resources are not restricted to

material commodities, but also include such things as friendship, recognition, health, or power. Values are goal states, or conceptions about the *desirable* level of goods. Lasswell's conceptions have structured this understanding, and he asserts that eight categories will exhaustively classify both goods and values. In his schema, there are four deference values (power, rectitude, respect, and affection) and four welfare values (wealth, well-being, enlightenment, and skill) (Lasswell & Kaplan, 1950:55).

To assess changes in value priorities over time, American Republican and Democratic party platforms from 1844 through 1964 were content analyzed, using procedures described by Stone, Dunphy, Smith, and Ogilvie (1966).<sup>1</sup> The use of content analysis is predicated by two assumptions: (1) The differential occurrence of a content category is an indication of the differential concern with the value

\*From pages 649–664, 671–677, and 680–683, with the permission of the editors of *The Journal of Interdisciplinary History* and The MIT Press, Cambridge, Massachusetts. © 1973 by the Massachusetts Institute of Technology and The Journal of Interdisciplinary History, Inc.

classified by that category; and (2) the relative value concern which is thus measured is an appropriate measure of the relative priority of that value in the total value schema of each and all documents. The content analysis, then, produces a profile of frequency changes in reference to seventy-three categories.<sup>2</sup> What are these categories?

About 95% of the words which occur in party platforms were entered in a dictionary, and these words were defined by one or more of the seventy-three categories. Many of these are really subcategories of the Lasswell Value categories. When possible, a distinction was made between categories (and, therefore, words) which indicate a substantive

**Table 1** Classification of the Value Dictionary

<b>I. Deference Values</b>	<b>Substantive Values</b>	<b>Value Transactions</b>
1. Power	other authoritative power cooperation solidarity conflict doctrine	Arena indulgence deprivation scope value indicator general participant authoritative participant
2. Rectitude	ethics religious	indulgence deprivation scope value indicator
3. Respect	other	indulgence deprivation
4. Affection	other	deprivation participant
<b>II. Welfare Values</b>	<b>Substantive Values</b>	<b>Value Transactions</b>
1. Wealth	other	transaction participants
2. Well-being	somatic psychic	indulgence deprivation
3. Enlightenment	other	indulgence deprivation scope indicator
4. Skill	aesthetics other	participants
<b>III. General Value Transaction Indicators</b>		
1. transaction indulgence		
2. transaction deprivation		
3. transaction		
4. scope indicators		
5. base indicators		
6. arena		
7. participant		
8. nations		
9. self		
10. audience		
11. others		
12. selves		
<b>IV. Anomie</b>		
1. anomie		
<b>V. Sentiments</b>		
1. positive affect		
2. negative affect		
3. not		
4. sure		
5. if		
<b>VI. Space-Time Dimension</b>		
1. space-time		
<b>VII. Residual Categories</b>		
1. n-type word		
2. undefinable		
3. undefined		

value concern and categories which indicate a concern with some value transaction whereby the actor (participant) gains (indulgences) or loses (deprivations) in a particular value environment (arenas, countries, etc.). Also, some categories indicate whether the values and the considerations are intrinsic for the participant (scope values) or instrumental (base values). If a subclassification were not feasible, words were classified in a residual category (other). For some words it is unclear what the particular value reference is; they are classified as general value unspecific indicators.

Three residual categories deserve further explanation: *n-type* words are high frequency words with little semantic information, such as articles and conjunctions. The category *undefinable* contains words that have no value implications whatsoever. The category *undefined* includes words with ambiguous value implications, which will change from context to context.

Armed with this instrument and using computers, the full text of the platforms was matched with the dictionary (or word classifications) and this matching produced the noted frequency profiles. Even if one were to agree that these frequencies may indicate changing value preferences in party platforms, the reader may well question the relevance of such data. Why bother with party platforms?<sup>3</sup>

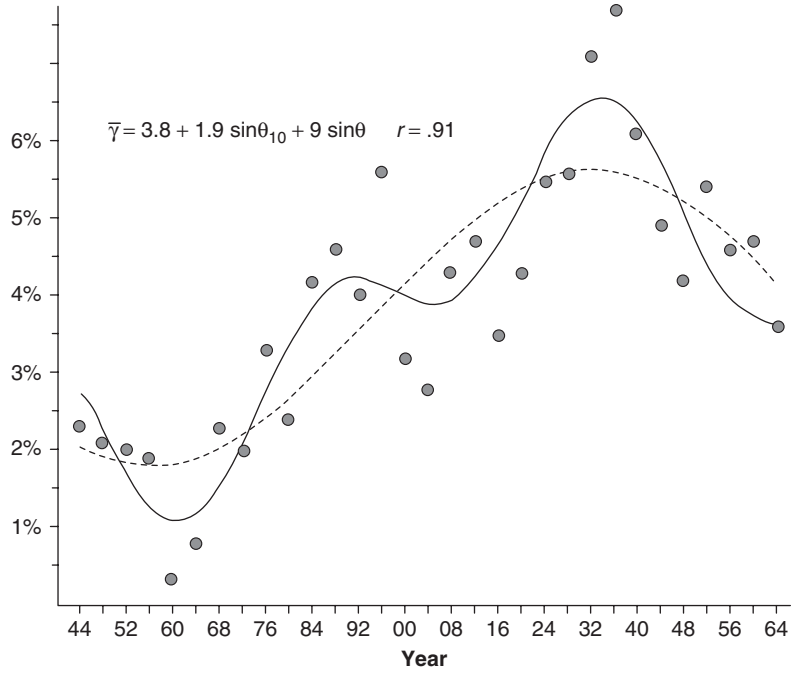
The choice of party platforms to assess magnitude and direction of changing values in American society seems justified for the following reasons: (1) The two-party system in the United States is competitive in most states of the Union, i.e., the parties compete in the same electoral market for the sympathies of various interests. The planks therefore contain the platform committee's best guesses about policies and values that will maximize the party's appeal to the electorate, and, in order to survive, parties must guess their voters' preferences correctly more often than not. Consequently, the content of party platforms is especially suitable for the study of values of the whole society. (2) Party platforms not only reflect predominant values, but they also create or modify value orientations by their

presentation and the ensuing public disputes during election campaigns. (3) Parties and party platforms are features of many other societies so that their examination allows for future cross-national comparisons.

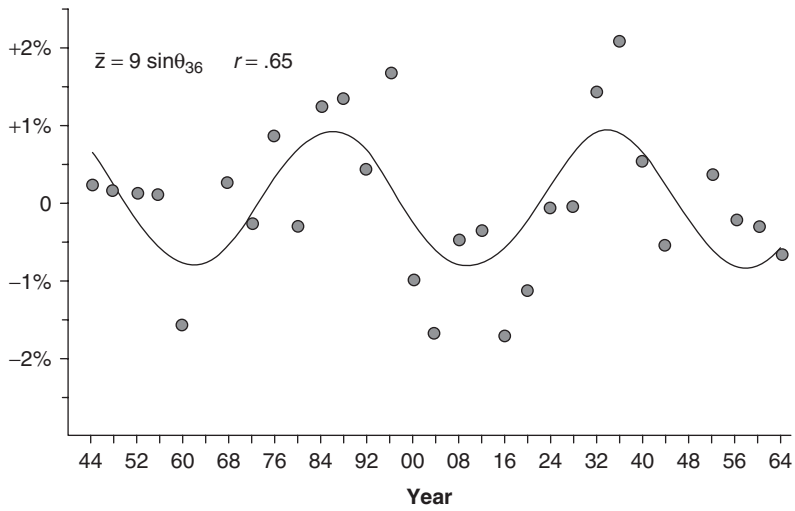
## DATA AND CYCLES

Basic data of this investigation are as follows: For each Democratic and Republican platform and for each campaign from 1844 to 1964 (or thirty-one campaigns), there are seventy-three observations, one observation for each category (or variable). Each observation is the frequency of that category in the particular platform. This frequency is then expressed as a percentage of words in that category of all words in the document, since this manipulation controls for the fact that campaign documents are of varying lengths. A plot of the thirty-one observations for the category "wealth-total" (a summary measure of all wealth subcategories) over the years 1844–1964 indicates that the concern with wealth varies a good deal from campaign to campaign.

Figure 1 illustrates . . . that in general the concern with *wealth* is low in the 1840s and 1850s; it increases over the next eighty years, to decrease again after 1932. This long-term cyclical tendency is estimated by a sine curve (the dotted line . . .). As will be noted, the actual observations do not lie on the dotted line. . . . If we plot the deviations from the dotted line (residuals) over time (see Figure 2, which represents these deviations for the Democratic platform), then we note a secondary cyclical trend which in the nature of the case has a more limited swing (or amplitude) and a shorter time span. This secondary cycle is also described by a sine curve which varies about the first one, and these secondary curves are represented by the drawn line in Figure 1. In conclusion, two cyclical trends seem to describe, if not operate on, changing concern with *wealth* in American platforms. . . . [S]imilar cycles tend to operate in most other value categories as well. How did I arrive at the latter conclusion?



**Figure 1** Two Superimposed Sine Curves Fitting Percent Concern With References to Wealth (Wealth Total) in Democratic Party Platforms, 1844–1964



**Figure 2** Short-Term Sine Curve Fitting Deviations From Long-Term Cycle Describing References to Wealth in Democratic Party Platforms, 1844–1964

Plots of the data revealed provisional outlines of a curve in each category and therefore the amplitude, wavelength, and year of maximum (or minimum) of each of these curves. These first estimates were subsequently tested and adjusted by an iterative computer program. A particular sine curve is considered an acceptable estimate of the underlying cyclical trend if it correlates with the data at  $r = .45$  or better (Namenwirth & Ploch, 1968; Porter & Johnson, 1961). This conservative decision rule is not wholly arbitrary since it provided a unique solution only in these cases. In a similar manner, if a short-term cycle correlated  $.40$  with the data, I accepted the existence of a secondary curve. Of the forty-two categories, about 80% displayed some type of cycle.

To state that sine curves approximate a good part of value change is not just to say that value changes display fluctuations, but that they display fluctuations of a particular kind. First, values fluctuate around an average level of concern, which is constant over time. Second, the magnitude of these fluctuations (or amplitude) is also constant over time. Third, the time span of each wavelength is constant, as well. In the case of the primary curve, the data, therefore, suggest static equilibrium. In the case of the secondary curve, the findings suggest a moving equilibrium, since the curve varies about an average level, which itself, is subject to constant change i.e., the primary curve.

In this manner, one can conceive of all value concerns and their changes as consisting of long-term curves, short-term curves, and detrended fluctuations. These three component parts have their own causes and dynamics.

#### THE FIT OF LONG-TERM VALUE CYCLES

Table 2 presents all of the content categories which fit a longer-term sine curve and four characteristics of each curve: (a) the party, i.e. Democratic or Republican; (b) wave length (or time span) in number of years; (c) the peak

(or year when the curve is at its maximum); (d)  $r^2$ , a measure of goodness of fit. The table indicates that in the long run, concern with the category *others* in the Republican platform, for instance, is at its height in the year 1808, and that this will again be the case in the year 2040 (i.e.,  $1808 + 232$ ). The long-term cycle explains about 64% of the variance in changing references to the category. In other words, the curve fairly well describes the varying usage of this concept over time, since the variance explained would equal 100% if it were to describe the variation perfectly, and zero percent if it were not to describe this variation at all. Even so, there was no party platform in 1808, and the statement about the platform in 2040 is only a projection into the future. The estimation of long-term change in value concerns is therefore often based on extrapolation.

In the Republican platforms, about three-fifths of the content categories, and more than seven-tenths of the categories in the Democratic platforms, were estimated by the long-term sine curve. Approximately one-fourth of the categories did not fit a long-term sine curve in either the Democratic or Republican Party platforms.

Unfortunately, the fifty-five long-term sine curves do not all have the same wavelength, and this complicates their interpretation. Although the modal wavelength is 152 years, the shortest cycle runs 104 years, while the longest lasts 232 years—or more than twice as long. Table 3 presents the specifics.

#### THE STRUCTURAL INTERDEPENDENCE OF LONG-TERM VALUE CHANGE

The internal relationships among the long-term sine curves are presented by a circle. To the right are the sine curves, which peak in subsequent years; to the left, are the curves, which peaked in previous years. The whole circle represents a 152-year sequence of peaking and dropping concerns with a variety of values. For instance, while long-term concern with the category *wealth* peaks around 1932, long-term concern with the categories *affection*, *respect*,

**Table 2** Selected Characteristics of 55 Long-Term Sine Curves

<i>Content Category &amp; Party<sup>a</sup></i>	<i>Peak</i>	<i>Wave Length (in Years)</i>	<i>r<sup>2</sup></i>
1. <i>Others</i>	1808	232	0.64
2. <i>Undefined</i>	1816	184	0.46
3. <i>Rectitude Scope Indicator</i>	1820	232	0.37
4. <i>Rectitude Scope Indicator</i>	1860	168	0.43
5. <i>Respect Indulgence</i>	1864	168	0.46
6. <i>Rectitude Total</i>	1864	152	0.40
7. <i>Affection Total</i>	1864	152	0.30
8. <i>Respect Total</i>	1868	152	0.49
9. <i>Rectitude Total</i>	1868	136	0.45
10. <i>Affection Total</i>	1872	120	0.38
11. <i>Respect Total</i>	1872	152	0.21
12. <i>Respect Indulgence</i>	1880	136	0.32
13. <i>Power Authoritative Participant</i>	1880	184	0.28
14. <i>Power Participant</i>	1884	136	0.22
15. <i>Power Authoritative</i>	1888	168	0.35
16. <i>Power Authoritative Participant</i>	1896	152	0.18
17. <i>Positive Affect</i>	1924	152	0.62
18. <i>Wealth Participant</i>	1928	136	0.58
19. <i>Skill Total</i>	1928	184	0.53
20. <i>Wealth Transaction</i>	1928	184	0.34
21. <i>Nations</i>	1928	184	0.24
22. <i>Wealth Total</i>	1932	152	0.73
23. <i>Wealth Other</i>	1932	152	0.73
24. <i>Wealth Other</i>	1932	152	0.69
25. <i>Wealth Total</i>	1932	152	0.69
26. <i>Skill Other</i>	1932	184	0.51
27. <i>Wealth Participant</i>	1932	136	0.46
28. <i>Wealth Transaction</i>	1936	152	0.32
29. <i>Transaction Indulgence</i>	1940	184	0.54
30. <i>skill Total</i>	1944	152	0.67
31. <i>Selves</i>	1944	168	0.65
32. <i>Selves</i>	1944	184	0.30
33. <i>Skill Other</i>	1948	152	0.69
34. <i>Well-being Total</i>	1948	136	0.34
35. <i>Transaction</i>	1948	168	0.26
36. <i>Transaction</i>	1948	184	0.25
37. <i>Well-being Total</i>	1952	136	0.68
38. <i>Well-being Somatic</i>	1952	152	0.32
39. <i>Well-being Somatic</i>	1956	152	0.74
40. <i>Base Indicator</i>	1956	168	0.70
41. <i>Transaction Indulgence</i>	1956	184	0.65

a. *Italics* = Republican Party Platforms; otherwise, Democratic Platforms.





Party platforms, such references often stand for a concern with the “other” party, but more often with people in general, and their wishes and qualities. A frequent usage of this category invokes a distinction between leadership and the masses, the “they” and all of those without a name, and it therefore indicates an elitist orientation to social reality and the political order.

The category affection contains references to love and friendship in general, and in party platforms, such references often indicate an association between devotion to family life and loyal patriotism. For instance;

Resolved that, with our Republican Fathers [affection-participant], we hold it to be a self-evident truth, that all men are endowed with the inalienable right to life, liberty, and the pursuit of happiness. . . . (Porter & Johnson, 1961:27)

In party platforms, the category *respect* includes the words honor, equality, and inequality. The category *rectitude* contains recurrent words like ought and must, which suggest a call for natural duty and principles. To illustrate:

We recognize the equality [*respect-others*] of all men before the law and hold that it is the duty [*rectitude-scope-indicator*] of the government in its dealings with the people to mete out equal and exact justice [*rectitude-ethics*] to all, of whatever nativity, race, color or persuasion, religion [*rectitude-religious*] or politics. (Porter & Johnson, 1961:41)

Slavery was the preponderant issue in these years. Policy preferences on this score divided the parties and changed over time—the Democrats favored slavery and its extension into the territories, the Republicans opposed the latter. However, our findings pertain to similarities, not differences, between the two parties, and the mutual concern with *rectitude* indicates that whatever the nature of substantive policy differences, there was a great and similar concern with the justification of policy preferences. In addition, at that time the terms of justification were largely rectitudinal.

In the 1890s, concern with *rectitude*, *respect*, and *affection* declined while concern

with *power-authoritative participant* and *power-authoritative* was at a peak. These categories contain many words, but most frequent are references to the federal government and the constitution. At first sight, it seems as if the political issues remained the same, i.e., the relationship between the states and the federal government. However, the justification of policy preferences changed from ethical to legal grounds, from substantive to formal justice, from traditional to legal. To illustrate:

During all these years the Democratic party has resisted the tendency of selfish interest to the centralization of governmental [*power-authoritative*] power, and steadfastly maintained the integrity of the dual scheme of government [*power-authoritative participant*] established by the founders of this republic of republics. Under its guidance and teachings the great principle of local self-government has found its best expression in the maintenance of the right of the states and in its assertion of the necessity of confining the general government [*power-authoritative participant*] to the exercise of the powers granted by the constitution [*power-authoritative*] of the United States. (Porter & Johnson, 1961:97)

On further inquest, one notes that the central issue is no longer the relationship between federal and state government, but the role of the federal government in the creation and maintenance of the economic infrastructure of an industrial society. The parties are in conflict about this role in regard to tariffs, transportation, politics (domestic as well as international, i.e., a canal through the isthmus), homesteading, banking and monetary policy, antitrust legislation, immigration, and taxation. Yet, the essential conclusion remains the same. Divergent policy preferences may appear from party to party and from campaign to campaign, and the justification of the divergent preferences is in terms of very identical legalistic constructs.

By 1932, the role of the federal government is much less disputed. The party program does not elaborate on justification, but simply states its preferences in regard to economic policy. The peaking concern with *wealth* and its subcategories indicates this

finding. Thus, one reads in the Democratic platform:

We favor the maintenance of national credit [*wealth-other*] by a federal budget [*wealth-other*] annually balanced based on accurate executive estimates within revenues [*wealth-other*], raised by a system of taxation [*wealth-other*], levied on the principle of ability to pay [*wealth-transaction*]. We advocate a sound currency [*wealth-other*] to be preserved at all hazards and an international monetary [*wealth-other*] conference called on the invitation of our government to consider the rehabilitation of silver [*wealth-other*] and related questions. (Porter & Johnson, 1961:331)

And the Republican platform states:

Generally in economic [*wealth-other*] matters we pledge the Republican Party: 1. to maintain unimpaired the national credit [*wealth-other*]. 2. To defend and preserve a sound currency [*wealth-other*] and an honest dollar [*wealth-other*]. 3. To stand steadfastly by the principle of a balanced budget [*wealth-other*]. . . . (Porter & Johnson, 1961:350)

First, there is a near total absence of justification of policy preferences. Second, a preoccupation with the material and technological (*skill*) well-being of the nation is a characteristic for the platforms at this time. This is further confirmed by the frequent references to the category *selves*. In party platforms, the use of we, us, ourselves, etc., often reveals a denial of status differentiation, either within the party or within the nation as a whole.

Typical for platforms in the 1950s and 1960s is an orientation toward the future rather than the past. This is revealed by maximum preoccupation with the category *transaction*.

We shall [*transaction*] insist [*transaction*] on businesslike and efficient administration of all foreign aid . . . We shall [*transaction*] erect [*transaction*] our foreign policy on the basis of friendly firmness. . . . We shall [*transaction*] pursue [*transaction*] a consistent foreign policy. . . . We shall [*transaction*] protect the future. . . . (Porter & Johnson, 1961:453)

Frequent references to *base* and *scope indicators* with words such as plan, strategy, future, project, and development, point in the same direction. In this framework, there is also maximum concern with health and *well-being* in general.

Projecting the findings into the future, the major concerns of the 1970s will be again of a different order. Since the 1930s, society as a whole was the object of concern; in the 1970s, the major preoccupation will be with conflicting groups and individuals. Frequent words in the subcategory *power-cooperation* are agreement, coalition, compromise, cooperative, organization, solidarity, unity, and, in the subcategory *power-conflict*, agitation, anarchy, breakdown, disagreement, disunity, fight, hostility, rebellion, resistance, and revolution. Seemingly, problems of the distribution of power and other social resources will be the major value issue at that time.

#### THE FIT OF THE SHORT-TERM VALUE CYCLES

Table 4 presents all of the categories that fit a short-term sine curve. According to this table, Republican concern with the category *well-being-total* tended to be at its height in 1908, as it did in 1868 and 1948. The short-term sine curve explains 26% of the residual variance in the category over time. The shorter-term cycle is therefore only a tendency in the data, which often explains but a limited part of the total amount of value change. In addition, the estimation of the cycles is based in part on extrapolation.

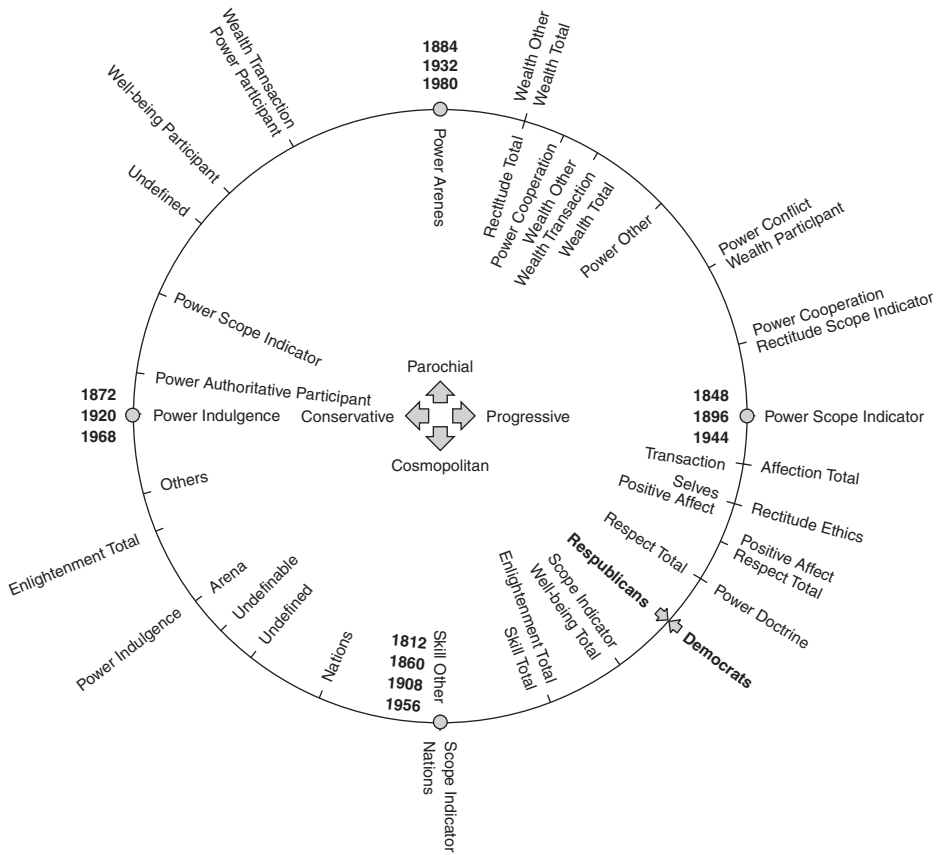
As was done with the long-term cycles, the wavelengths of the shorter-term cycles were set at the modal length of forty-eight years and peaks transformed accordingly. In this case, the wavelength of the transformed sine curve is considerably less than the period under observation, and, therefore, each point on the circle represents a set of peaks, which are forty-eight years apart. For instance, the top of the circle (in Figure 4) represents the years 1884, 1932, and 1980.

**Table 4** Selected Characteristics of Short-Term Sine Curves

<i>Content Category and Party<sup>a</sup></i>	<i>Peak</i>	<i>Wave Length (in years)</i>	<i>r<sup>2</sup></i>	<i>% Total<sup>b</sup> Variance</i>
1. <i>Well-being Total</i>	1908	40	0.26	0.17
2. <i>Arena</i>	1908	68	0.22	—
3. <i>Respect Total</i>	1908	36	0.19	0.1
4. <i>Respect Total</i>	1910	32	0.22	0.18
5. <i>Others</i>	1912	68	0.23	0.08
6. <i>Enlightenment Total</i>	1916	52	0.21	0.15
7. <i>Power Indulgence</i>	1916	44	0.2	0.11
8. <i>Scope Indicator</i>	1916	32	0.17	0.03
9. <i>Undefined</i>	1920	80	0.3	0.16
10. <i>Power Authoritative Participant</i>	1920	52	0.23	0.19
11. <i>Undefinable</i>	1920	32	0.2	—
12. <i>Power Scope Indicator</i>	1926	32	0.3	—
13. <i>Power Indulgence</i>	1926	32	0.17	0.13
14. <i>Well-being Participant</i>	1928	32	0.25	0.14
15. <i>Wealth Transaction</i>	1928	48	0.17	0.13
16. <i>Power Participant</i>	1928	48	0.12	0.09
17. <i>Power Arena</i>	1932	40	0.24	—
18. <i>Wealth Other</i>	1934	48	0.43	0.13
19. <i>Wealth Total</i>	1934	48	0.43	0.13
20. <i>Rectitude Total</i>	1934	68	0.35	0.19
21. <i>Wealth Total</i>	1936	48	0.26	0.07
22. <i>Wealth Other</i>	1936	48	0.24	0.06
23. <i>Wealth Transaction</i>	1936	48	0.2	0.15
24. <i>Power Cooperation</i>	1936	64	0.18	0.15
25. <i>Power Conflict</i>	1938	36	0.25	0.19
26. <i>Power Cooperation</i>	1940	40	0.33	0.18
27. <i>Power Other</i>	1940	64	0.3	—
28. <i>Wealth Participant</i>	1940	48	0.21	0.11
29. <i>Rectitude Scope Indicator</i>	1942	48	0.24	0.15
30. <i>Undefined</i>	1944	20	0.21	—
31. <i>Transaction</i>	1944	44	0.17	0.13
32. <i>Selves</i>	1946	48	0.27	0.19
33. <i>Affection Total</i>	1946	52	0.23	0.16
34. <i>Positive Affect</i>	1946	44	0.17	0.06
35. <i>Power Scope Indicator</i>	1948	64	0.23	—

a. *Italics = Republican Party Platforms; Otherwise, Democratic Platforms.*

b. Where blank, curve fitted to raw data.



**Figure 4** The Internal Structure of Short-Term Value Changes (Cycle Length Set at 48 Years, Variable Origins)

**AN EXPLANATION OF SHORT-TERM CYCLICAL VALUE CHANGES**

The reader will have noted that the early 1890s and the years around 1932 represent periods of sustained depression and business contraction in the American and world economy. In addition, twenty-five to thirty years earlier or later were periods of sustained economic growth (Burns & Mitchell, 1946:429; Fellner, 1956:43–54; Gordon, 1952:235–243). In short, there appears a rather striking fit between the short-term wheel of time and a particular economic cycle. What is the latter cycle?

Economists distinguish between various cyclical fluctuations in business activity: seasonal fluctuations, the business cycle, the long wave, and the secular trend (Fellner,

1956). The long wave, even though disputed by some economists, is said to vary, extending and contracting over a period of fifty to sixty years. How does the latter process relate to value articulations?

During long-wave economic deterioration, the nation turns inward, gradually relinquishing international ventures and then obligations, becoming more and more parochial in its orientations. This parochialism is first conservative, probably stressing discipline, the tightening of belts, the necessity of temporary unemployment as well as charity to overcome the economic decline. Usually, this goes together with growing indifference, if not hostility, toward foreign claims and conditions as the outside world will be seen as competitive, fickle, a cause of troubles, and an object of scapegoating. With the ongoing but

diminishing calamity, the mood will change from conservative to progressive. Increasingly, belt tightening and charity will be seen as palliatives. A growing demand will arise for a change in collective arrangements and structural intervention. Whether cause or effect, the ensuing structural change seems to work since prosperity returns. With increasing surplus, attention turns again to the world scene; value articulations become more cosmopolitan, at first in a progressive vein. Progressive intervention works at home; therefore, it needs to be exported in the fulfillment of America's ethos and liberal designs. At any rate, money is there in growing abundance. However, once the expansion turns its peak and contraction sets in, the cosmopolitan impulse turns from progressive to conservative, from national mission to national interest, from, for instance, Marshall Plan to Green Berets. One may well speculate that with Vietnamization and the Nixon Doctrine, the parochial phase is on the rise again.

The relationship between long-wave economic contraction and expansion, and a shift in basic understandings regarding the nature of morality and criteria of worth, are equally systematic. Briefly, the shift from conflict to consensus and *vice versa* commences at the beginnings of the period of sustained debacle and prosperity, while the shift from particularistic to universalistic and *vice versa* begins at the onsets of sustained growth and contraction.<sup>4</sup>

### CONCLUSIONS

A content analysis of American party platforms produced results, which seem to fit a variety of trends in a great many different value categories. In addition to a long-term trend of about 148 years, one can often discern short-term trends of about forty-eight years. The latter represent variations in value concern over and above long-term trend variations. In combination, the short- and long-term cycles describe (or explain) a good part of the variation in value concerns.

The wheels of time summarize the internal structure of both types of value cycles.

Sequentially, the varying value concerns of the long wheel of time are explained in terms of four fundamental functional problems of any society. Accordingly, the solution of one problem always takes precedence over the solution of the next one until all four problems—adaptive, instrumental, integrative, and expressive—have been articulated to the fullest and the progression commences anew.

The sequential articulations of the short wheel of time—parochial, progressive, cosmopolitan, and conservative—are most likely produced by a dynamics that differs from the long-term functional mechanisms, and the “long wave” periodic contraction and expansion of the national economy seems, for the moment, the most plausible explanation.

Long- and short-term dynamics are not equally important in the determination of value change. On the average, long-term cycles describe about three times as much of the variance in value change as do short-term cycles. The larger part of changing value articulations in platforms is therefore attributed to the dynamics of social problem-solving rather than to social structural changes. Yet, the theory is not purely functional, since it is suggested that economic mechanisms are operating beyond and above the functional dynamics.

In the exposition, it is assumed that the time span and magnitude of value change are constant for all times. This seems an unwarranted assumption. Indeed, if the “long wave” explains the shorter-term wheel of time, then the sine curve may well be too constrictive a model of value change. Even though the “long wave” is a recurrent and rhythmically alternative cycle, the magnitude and wavelength of these cycles seem to vary in history. If this is the case for the cause, so it must be for the consequences, and thus the sequence of political philosophies must be of varying duration. One would like to believe that changes in duration and amplitudes are themselves a simple function of time and therefore gradual and continuous, but the world of value transformations may not submit itself so readily to this persistent search for elegance and order.

Quantitative procedures, such as content analysis and curve fitting, suggest to the

uninitiated reader an exactness and precision which are far greater than the results of more customary procedures of historical analysis. This practitioner is under no such illusions. I forewarned the reader about the approximate nature of estimates and the speculative character of subsequent interpretations and explanations. Their correctness cannot be established in one experiment, and judgments on that score must await future examinations of different historical sources using different procedures of analysis.

#### NOTES

1. In 1844, 1848, and 1852, there were no Republican platforms since the Republican Party did not exist prior to 1856. For the first three campaigns, I used the Whig platforms because the Whigs are in many respects the precursors of the Republican Party.

2. Thirty-one categories were eliminated from the analysis because of low frequencies or poor distribution. For a discussion of all of the categories, see Namenwirth and Weber, 1987.

3. The study of party platforms has other uses besides the assessment of value changes. See Benson, L. (1961). *The concept of Jacksonian democracy: New York as a test case* (Princeton, N.J.: Princeton University Press), p. 216, and Klingberg, F. L. (1952). The historical alternations of moods in American foreign policy. *World Politics*, IV:239–274.

4. For a general discussion of the relationship between economic cycles and political if not philosophical thought, see Pareto, V. (1935). *The mind and society*. New York: Harcourt & Brace, paragraph 2387.

#### REFERENCES

- Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. New York: National Bureau of Economic Research.
- Fellner, W. J. (1956). *Trends and cycles in economic activity: An introduction to problems of economic growth*. New York: Holt.
- Gordon, R. A. (1952). *Business fluctuations*. New York: Harper.
- Lasswell, H. D., & Kaplan, A. (1950). *Power and society: A framework for political inquiry*. New Haven, CT: Yale University Press.
- Namenwirth, J. Z., & Ploch, D. R. (1968). *Structural and contingent value changes in American political party platforms*. Unpublished master's thesis, Yale University, New Haven, CT.
- Namenwirth, J. Z., & Weber, R. P. (1987). *Dynamics of culture*. Boston: Allen and Unwin.
- Porter, K. H., & Johnson, D. B. (1961). *National party platforms, 1840–1960*. Urbana: University of Illinois Press.
- Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (Eds.). (1966). *The General Inquirer: A computer approach to content analysis*. Cambridge, MA: M.I.T. Press.

# 3.9

## INFERRING THE READABILITY OF TEXT

KLAUS KRIPPENDORFF

**W**e tend to attribute readability to written text. The 2000 *Oxford English Dictionary* defines it as “the quality of, or capacity for being read with pleasure or interest, considered as measured by certain assessable features as ease of comprehension, attractiveness of subject and style.” This definition backgrounds the obvious—that it requires a reader for texts to be readable, in fact, for texts to be texts. Readability, like literacy, is a cultural phenomenon, and efforts to infer readability from text serve various social institutions, in effect creating functional differentiations of individuals’ ability. Readability research provides content analysts with an interesting case study of how analytical constructs are constructed and applied.

### INFERRING READABILITY FROM TEXTS: A BRIEF HISTORY

Educational research pioneered efforts to measure readability in the 1920s. Readability became an issue in deciding on reading material appropriate for schoolchildren on different levels. In the 1930s, readability research

expanded to adults, serving the emerging needs of industry, government, and the military to evaluate magazines and books for their publishability, as well as technical communications, training manuals, and forms for their reliable use in processes of an administrative nature. Now, leading word-processing software features readability measures, intended to aid good writing.

The guiding idea of readability research is to find an index of readability that is general, not content specific. This is why traditional content analysts have not participated in its development. However, what students of readability and content analysts have in common is the effort of making reliable and valid inferences from text to a chosen context, as well as the need to connect the two empirical domains by means of what content analysts call analytical constructs. The whole history of readability research is one of gradually refining definitions of readability so that it can be inferred from measurable textual attributes and of improving the underlying analytical constructs.

Early readability studies looked into vocabularies—cataloging words with which

students on various grade levels would be familiar (Thorndike, 1921) and could be expected to cause few reading problems. In the 1930s, readability researchers began to employ statistical correlations between numerous measurable attributes of texts and judgments of the difficulty of reading these texts. Regression equations were used to identify predictors of readability from which computationally efficient formulas could be constructed. For schoolchildren, a series of test lessons, developed by McCall and Crabbs (1925), still published, increasingly became the standard criterion against which most readability measures were tested. For adults, researchers used the opinions of library users, the popularity of publications, and multiple-choice comprehension tests. Adults showed more diversity than did children owing to the influence of different interests, backgrounds, levels of education, and occupation, rendering a common formula more difficult.

In 1934, Rudolph Flesch, probably the most cited readability researcher, proposed a simple three-factor formula, which correlated .74 with McCall-Crabbs test scores (Klare, 1963:56ff). His procedure:

*Systematically select* samples of 100 words throughout the material to be rated

*Compute* average sentence length in words ( $x_s$ )

*Count* the number of affixes ( $x_m$ )

*Count* the number of personal references ( $x_h$ )

*Average* the results and insert in the formula:

$$.1338x_s + .0645x_m - .0659x_h - .7502.$$

The resulting index brought most of the measured texts between 1 (easiest) and 7 (most difficult). Adding the constant 4.2498 instead of .7502 gave the reading grade placement at which 75% comprehension could be observed. His formula acknowledged earlier findings that long sentences are difficult but added the intuition that abstractions, indicated by affixes, add to this difficulty while personal references subtract from it.

An interesting controversy led Flesch to modifications of his formula. According to

Klare (1963:58), the statistician S. S. Stevens, known for his distinctions of four levels of measurement, and Geraldine Stone (1947) applied the Flesch formula to psychology textbooks used at Harvard University and found the difficulty of William James's *Psychology* to be overestimated and Koffka's *Principles of Gestalt Psychology* to be underestimated according to student judgment. Following this controversy, Flesch (1948) separated readability into two kinds: reading ease and human interest. Because of the laborious nature of counting affixes, in his new reading ease formula, the count of affixes was replaced by a count of the number of syllables per 100 words. Flesch's reading ease (see below) continued to correlate highly with the McCall-Crabbs criterion, .7047, whereas human interest was .4306. Counting syllables has been the most common feature of most readability measures ever since. Flesch did not abandon his earlier insight that reading ease had much to do with using abstractions and developed a measure of the level of abstraction (Flesch, 1950) of words in a text. This could be used as such but also as a readability measure because it correlated .72 with the criterion, a slight increase over the .7047 for the reading ease measure.

The 50 years following Flesch's and his predecessors' proposals were filled with readability studies. Some 40 formulas have been proposed, tried out, abandoned, or refined—not all of them tested by empirical evidence (DuBay, 2004). But the analytical constructs evolved very little.

## CURRENTLY POPULAR ANALYTICAL CONSTRUCTS

### Flesch Reading Ease (Flesch, 1948)

This formula has now withstood the test of time. It calls for selecting any 100-word sample from a text, counting

ASL = the average sentence length = total number of words/total number of sentences

ASW = the average number of syllables per word



and computing

$$206.835 - 1.015 \text{ ASL} - 84.6 \text{ ASW.}$$

The resulting score ranges from 0 to 100 and suggests that fifth graders can read texts at score 90 to 100, eighth to ninth graders at 60 to 70, and college graduates at 0 to 30. According to Wikipedia, *Readers Digest* magazine scores about 65, *Time* magazine about 52, and *Harvard Law Review* below 30. Microsoft Word computes this score after using its spell checker. This chapter scores 29.9.

### Flesch-Kincaid Grade Level

This formula translates the components of Flesch's reading ease into a score that reflects grade levels of education, making it easier for teachers, parents, and librarians to recommend appropriate reading materials, including books, to students. It too calls on counting:

ASL = the average sentence length = total number of words/total number of sentences

ASW = the average number of syllables per word

but computing this index:

$$0.39 \text{ ASL} + 11.8 \text{ ASW} - 15.59.$$

Its score corresponds to the grade level at which an average student would be able to read the measured text. Microsoft Word produces this automatically as well. This chapter scores 12.6, which would suggest a level below college.

### Passive Sentence Readability

Microsoft Word also computes the proportion of passive to all sentences of a document. It ranges from 0 (supposedly easiest) to 1 (supposedly most difficult) and is based on the contention that passive sentences weaken the direction of the verb and can confuse the meaning of a sentence, even when grammatically correct. A reference for this index could not be found, nor evidence of its validity. Since it has become available, it is being

discussed. Writers insist that one cannot do entirely without passive constructions, suggesting that 25% would still be readable. In this chapter, 9% of sentences are passive.

### Dale and Chall Formula (Dale & Chall, 1948)

These researchers sought to improve on Flesch's AWS, replacing it by a count of difficult words, defined as not occurring on a carefully researched list of 3,000 easy words. Forty years later, Dale and O'Rourke (1981) revised this list. They recommend taking several 100-word samples from different parts of a text, for books every 10th page, counting

PDW = the proportion of words not on the list of 3,000 easy words = number of difficult word/total number of words

ASL = the average sentence length = total number of words/total number of sentences

and computing

$$0.1579 \text{ PDW} + 0.4996 \text{ ASL} + 3.6365.$$

Its score, designed to indicate the grade level, consistently correlated .70 with the McCall-Crabbs criterion.

### SMOG (Simple Measure of Gobbledygook) (McLaughlin, 1969)

SMOG scores too indicate reading levels, defined as the level at which readers can understand 90% to 100% of the information in a text. It calls for counting

NS = the number of sentences involved—at least 30: 10 consecutive sentences selected near the beginning of a text, 10 in the middle, and 10 near the end. In long sentences with colons or semicolons followed by a list, count each part of the list, together with the beginning phrase of the sentence, as an individual sentence.

NP = the number of polysyllable words in these sentences (i.e., words with three or more syllables, even if the same word appears more than once). Count words with hyphens as one. Read numbers aloud to determine the number of syllables it takes to verbalize them. Take abbreviations as the whole word they represent.

and computing

$$1.0430\sqrt{NP\left(\frac{30}{NS}\right)} + 3.1291.$$

**Fry Readability Formula (Fry, 1977)**

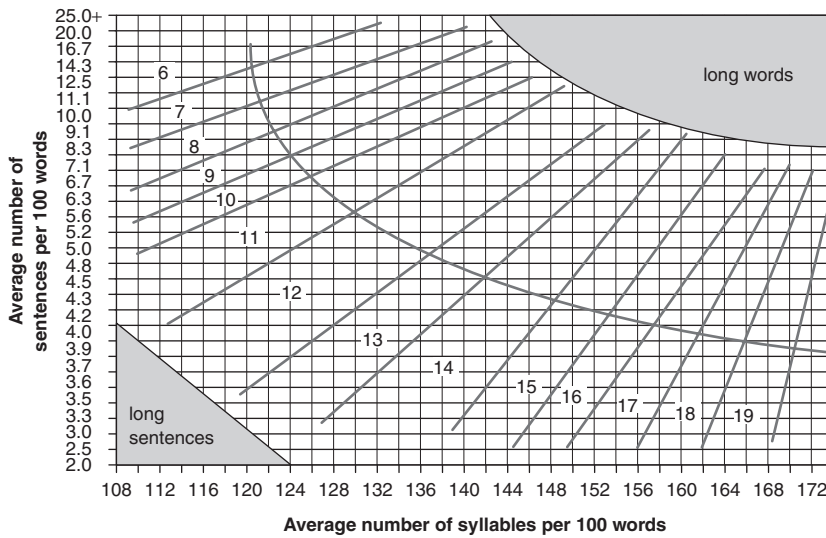
Fry's approach surely is most user-friendly. It requires randomly selecting three

passages of exactly 100 words, beginning with a sentence, counting

$X$  = the number of sentences in the 100 words, estimating the last sentence to the nearest 1/10th

$Y$  = the number of syllables in the 100 words

and finding the grade level in the intersection of the  $X$  and  $Y$  coordinates in the following graph.



**Figure 1** Fry Graph for Estimating Reading Ages (in Years)

**FORCAST Readability Formula**  
(Caylor, Sticht, Fox, & Ford, 1973;  
see DuBay, 2004, p. 51)

The use of this formula is even simpler than Fry's. It was developed to evaluate reading requirements in the U.S. Army, applied to military reading matters, especially technical instructions, and tested with members of the Army in various occupational roles. It asks to count

$N_1SW$  = the number of one-syllable words in a passage of exactly 150 words

and compute

$$20 - N_1SW/10.$$

This surprisingly simple formula was found to correlate .98 with Flesch's formula, .98 with Dahl-Chall's, and .77 with graded military reading matter. It had the advantage of working within a relatively homogeneous adult population of military recruits and service personnel. DuBay (2004:52) reports on a similar research project for the U.S. Navy.

**STRUCTURAL PROBLEMS  
OF THE PATHS TAKEN**

Over 80 years of efforts by a growing community of researchers to improve the validity of the construct underlying these formulae, multiple regression equations, have reached a ceiling. Correlations with readability criteria

seem to stay below .80. One can identify six reasons for this ceiling. They can serve as a warning to content analysts who seek to develop similar computational constructs for making inferences from text:

- *Surface measures.* Sentence lengths, numbers of syllables in words, short words, difficult words, and so on are easily countable but pertain only to epiphenomena of reading and writing. Long words, for example, are not difficult as such, but because they tend to be used less often, they naturally include more unfamiliar words than short words do. There are multisyllable words most English readers have no problems with, like *television*, and monosyllable words, like *hod*,\* that may well bring reading to a halt. Counts attend to the rarely noted surface of texts that readers typically penetrate.

- *Typography.* Webster's dictionary includes legibility in the definition of readability, originally good handwriting. As graphic artists know too well, the readability of printed matter is influenced by font styles, sizes, colors, and background, as well as by the organization of text (hierarchical organization of headlines, bullets, and highlighting devices) and layouts, including the use of illustrations and graphs, known to add to interest and comprehension.

- *Narrative structures.* Counts cannot capture the organization of the counted words, propositions, sentences, or paragraphs into larger compositions. Narratives, arguments, syllogisms, coherence, and the development of plots that writers consider crucial for making complex ideas clear escape context-free counting of units of text.

- *Readers' choices.* Readers of technical instructions, even of large newspapers, rarely feel bound to work through a text linearly, from its beginning to its end. They typically navigate through textual matter, selecting what supports the construction of

their own mental narratives along the way. Hypertext documents support nonlinearities explicitly.

- *Discursive competencies.* Students, during their formal education, constitute a population that is relatively easy to differentiate into grades, but adults develop unequal competencies and approach texts situationally. Discourse communities distinguish themselves by their members' interests, use of specialized vocabularies, customary patterns of reasoning, and prior knowledge of relevant subject matter. What is readable in one community may be incomprehensible in another. A single formula cannot do justice to this diversity.

- *Cultural dynamics.* Familiarity with vocabularies and grammatical constructions changes with reading experiences and over time. When used, what is difficult today is destined to be less so in the future. The difference in vocabulary between young people and older folks is not merely developmental, as supposed by formulae that predict reading grades. It signals a dynamics of culture and language use to which reading and writing contribute, literature and poetry in particular. For one example, in pre-Elizabethan English, the average sentence length was 50 words. In Elizabethan English, it was 45 words. In Victorian English, it was 29 words (Sherman, 1893, cited in DuBay, 2004). Currently, average sentence length is down to 20. As readability shifts, ways to infer it must do as well. Culture-free formulae cannot.

Writers and teachers associations have considered it a danger and resisted equating the components of readability formulae with guidelines for good writing. Indeed, if writers were rewarded by achieving high readability scores, and only that, they could easily produce meaningless strings of monosyllable words. This possibility suggests that the above formulae address only one epiphenomenon of reading and writing, not the heart

---

\*A tray with a pole handle that is borne over one's shoulder for carrying loads, typically mortar or bricks. Hods are most likely familiar to masons.

of it. Inferred readability is unlike actual readability. Readability has to do with how readers understand texts and whether their writers can put themselves in the place of readers and make sense of their texts: together. For writers, readability formulae can at best serve as a warning that something may have to be addressed.

### MEASURING THE READABILITY OF TEXTS: THE CLOZE PROCEDURE

The above formulae are analytical constructs, derived from multiple regressions of textual attributes on a criterion variable. They are used predictively, to infer not the readability of texts but various measures of it. Prediction takes place in advance of a text being read—other than by their authors. This inference implicitly assumes that the analytical construct underlying these formulae could model and represents readers in some respect. As shown above, this representation is shallow indeed.

Taylor (1953) developed a test that overcomes most of the above-mentioned shortcomings. It relies on an ability of readers that is most closely aligned with comprehension: their ability to anticipate and make sense of words from their context of use. Osgood (1959:78–88) introduced Taylor's "Cloze procedure" to the community of content analysts. The procedure is simple enough. From a text:

*Replace*, say, every fifth word by a blank that readers can fill in. Texts with at least 50 blanks make the procedure quite reliable.

*Count* the number of correct guesses of the deleted word—correct in form (no synonyms), number, person, tense, voice, and mode. Ignore differences in spelling.

The *proportion* of correct guesses is the Cloze score.

Cloze scores correlate highly with the multiple-choice answers by readers and can be considered a substitute for subjective judgments of reading difficulty.

Guessing the correct words from their context of use enables readers to employ most of the abilities that the above formulae must ignore: the information provided by typography, grammar, narrative structures, readers' discursive competencies, author-reader commonalities, and changes in culture—readers are always from the present and can be chosen from the population of interest.

Scores below .35 indicate frustration, .35 to .50 assisted reading, instructional, and .50 to .60 unassisted reading (DuBay, 2004:27). The Cloze procedure does not infer readability from textual attributes; it measures the redundancy needed for reading comprehension. It relies on real people rather than regression equations. Inferences are inductive (from a sample to a population of texts + readers), not abductive (from text to the human ability to read it).

### LESSONS FOR THE CONTENT ANALYST

For content analysts, the lessons from the history of readability research are as follows:

- It is important to be clear about the research questions to be answered and what is to be inferred, and to develop and empirically validate analytical constructs before applying them inferentially.
- To allow validity to accumulate. It pays to consider the categories and analytical constructs of previous research before inventing new categories whose validity is uncertain.
- Most analytical constructs have structural limitations. They reach ceilings in their ability to answer research questions, in their accuracy of the inferences they enable. Going beyond these ceilings requires structural innovations in the analytical constructs used to bridge the gap between text and what is to be inferred from it.
- Using the intellectual capabilities of readers, coders, and observers may introduce problems of reliability but can greatly enhance validity in the end.

## REFERENCES

- Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. (1973). *Methodologies for determining reading requirements of military occupational specialties* (Technical Report No. 73-5). Alexandria, VA: Human Resources Research Organization.
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin* 27:1-29, 37-54.
- Dale, E., & O'Rourke, J. (1981). *The living world vocabulary: A national vocabulary inventory*. Chicago: World Book-Childcraft International.
- DuBay, W. H. (2004). *The principles of readability*. Costa Mesa, CA: Impact Information. Retrieved October 25, 2006, from [www.impact-information.com/impactinfo/readability02.pdf](http://www.impact-information.com/impactinfo/readability02.pdf)
- Flesch, R. F. (1934). Estimating the comprehension difficulty of magazine articles. *Journal of General Psychology* 28:63-80.
- Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology* 32:221-233.
- Flesch, R. F. (1950). Measuring the level of abstraction. *Journal of Applied Psychology* 34:384-390.
- Fry, E. B. (1977). Fry's readability graph: Clarification, validity, and extension to level 17. *Journal of Reading* 24:242-252.
- Klare, G. R. (1963). *The measurement of readability*. Ames: Iowa State University Press.
- McCall, W. A., & Crabbs, L. M. (1925). *Standard test lessons in reading: Teacher's manual for all books*. New York: Bureau of Publications, Teachers College, Columbia University.
- McLaughlin, G. (1969). SMOG grading: A new readability formula. *Journal of Reading* 12:639-646.
- Osgood, C. E. (1959). The representational model and relevant research methods. In I. de Sola Pool (Ed.), *Trends in content analysis* (pp. 33-88). Urbana: University of Illinois Press.
- Sherman, A. L. (1893). *Analytics of literature: A manual for the objective study of English prose and poetry*. Boston: Ginn & Co.
- Stevens, S. S., & Stone, G. (1947). Psychological writing, easy at hand. *American Psychologist* 2:230-235.
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly* 30:415-433.
- Thorndike, E. L. (1921). *The teacher's word book*. New York: Bureau of Publications, Teachers College, Columbia University.