The Journal of Abnormal and Social Psychology
Vol. 48, No. 3, 1953

CODING NOISE IN A TASK-ORIENTED GROUP [1, 2]

JOSIAH MACY, JR., LEE S. CHRISTIE, AND R. DUNCAN LUCE
Massachusetts Institute of Technology

The existing experimental studies of small task-oriented groups have generally employed noise-free communication, in the sense of information theory (1, 2, 5). One exception is the work of Heise and Miller (4) in which measured amounts of acoustic noise were introduced into the telephone network connecting the members of a small group. The experiments reported here investigate the effects of semantic or coding noise on the performance of small task-oriented groups. In this case, the transmission of information along the transmission channel was noise free, but the coding and decoding processes were ambiguous. The ambiguity arose from the necessity for the subjects (*Ss*) to write and to interpret descriptions of colors that are not easy to describe. The term *coding noise* is used to refer to this ambiguity, as explained below in detail. For a more detailed discussion of semantic noise in relation to the general notions of information theory and for an analysis of parts of the experiment not discussed here, the reader is referred to Christie, Luce, and Macy (3).

Method

Apparatus

A round table partitioned into five *S* compartments, similar to that described by Leavitt (5), was used. From each compartment to every other one were slots sufficiently large to receive the 8 by 1½-inch message cards used by the *Ss*. Each *S* was identified throughout the experiment by the color of the cards on which he wrote his messages. The communication network to be studied was imposed by physically blocking the slots not to be used. The table differed from Leavitt's in that the answer signal switches were removed and for them were substituted rubber tubes running from each *S*'s compartment to *E*'s station.

Procedure

The *Ss* were told prior to entering the experimental room that at the start of each trial they would open a box containing five colored marbles; that only one color of marble would be in everyone's box; and that their task was to determine this color by written communication on cards sent through the allowed channels. When an *S* knew the answer, he was instructed that he was to drop the corresponding marble down the tube in his compartment. If he wished to change his answer before the end of a trial, he could drop a second marble as a correction but the trial would end when each of the five *Ss* had dropped at least one marble.

The *Ss* were then taken to the experimental room, seated at the table, and after a brief check to insure familiarity with the apparatus, the trials started. No talking was permitted after the trials began. The groups were run for 30 successive trials—the marbles used on the first 15 trials were drawn from a set of six different plain, solid colors, easy to distinguish and to describe. At the sixteenth trial and thereafter, the marbles used were from a set of six cloudy, mottled, indistinct colors. They were still easy to distinguish if they could be directly compared, but it was difficult to describe each one clearly and unambiguously.

The experimental program was divided into two parts, with a time lapse of about nine months between parts. In Part I, four groups of five *Ss*, volunteer M.I.T. undergraduates, were run on each of three networks—star, chain, and circle. In Part II, four groups of five *Ss*, enlisted military personnel from Fort Devens and the First Naval District Receiving Station, were run on each of three networks—circle, pinwheel, and star. (See Fig. 1 for diagrams of the networks.) Each *S* participated in only one experimental run. The experimental procedures in Parts I and II were as nearly identical as it was possible to make them, with the exception of the star groups of Part II. This will be discussed below.

Data Record

The data obtained consisted of a record of the marbles dropped by each man, the time to complete each trial, and the messages sent, sorted as to sender

[1] This work was supported in part by the Army Signal Corps, the Air Force Air Materiel Command, the Navy Department (Office of Naval Research), and the RAND Corporation.

[2] The experimental design and the results reported in Part I were carried out by A. Bavelas and S. L. Smith to investigate the adaptability of task-oriented groups to sudden changes in environmental conditions (Smith, S. L. Communication pattern and the adaptability of task-oriented groups: an experimental study. Mimeographed paper). The experiments reported in Part II and the analysis from the information theory point of view reported in this paper were carried out by the authors. The assistance of Patricia Thorlakson and Deborah Senft in the performance of the experiments and the calculations is gratefully acknowledged.

by color and as to receiver by position at the table to which they were sent.

Data on time for each trial and number of messages used showed the same relative differences between networks as previously reported for such experiments by Leavitt (5).

a trial with a wrong marble as his final answer, this was counted as one error.

Since only four groups were run in each network, and considerable variation in the error count was observed among groups run on the same network, it is important to know the confidence intervals for

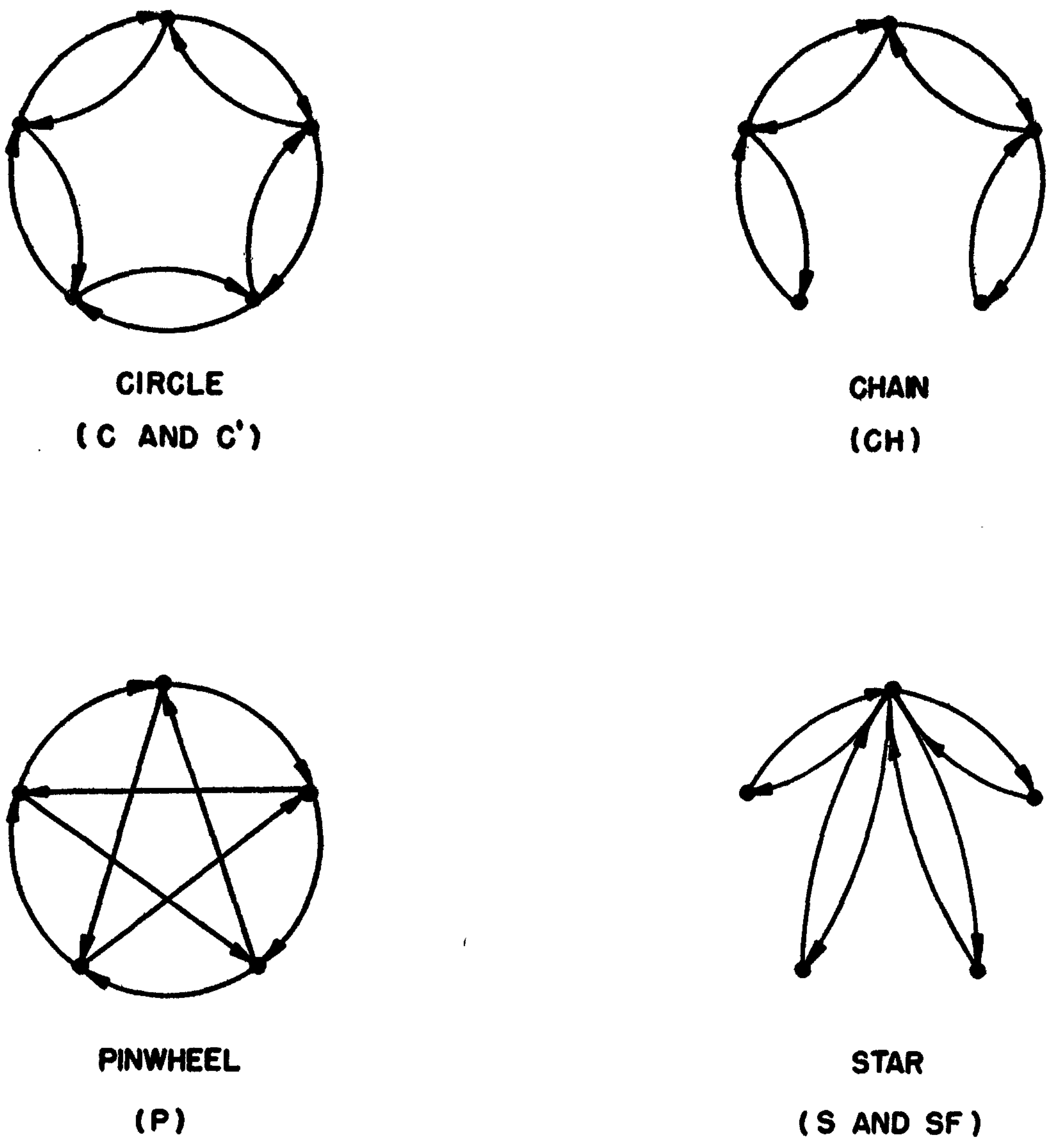

FIG. 1. COMMUNICATION NETWORKS USED IN PART I AND PART II

An examination of the raw data suggested that during the last 16 trials the groups had more difficulty obtaining the correct answer when certain colors were held in common than when other colors were held in common. The raw error data were therefore corrected so that all colors had the same relative error frequency, and these corrected errors are plotted in Figs. 2 and 3. Every time an *S* ended

these means. Figure 4 presents the .05 confidence limits, calculated from Student's *t* distribution. The rectangles in Fig. 4 represent the range within which the population mean may be expected to be in 95 per cent of the cases, so that those networks whose rectangles do not overlap for a given trial block are significantly different at better than the .05 level with regard to error count. All the groups had less than 10 per cent errors on all the first 15 trials, with no observable differences between networks.

Figs. 2, 3, and 4 show that though there are no significant differences in error count among net-

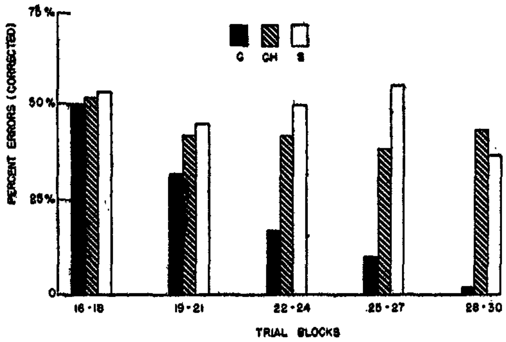

Fig. 2. Corrected Mean Percentage Errors for Experimental Groups in Part I

works in the first 15 trials, the introduction of the more difficult marbles at trial 16 greatly increased the errors made by all networks. Subsequently, certain networks learned to reduce their errors to the previous low value, while others did not. We will examine the mechanism of these effects in some detail.

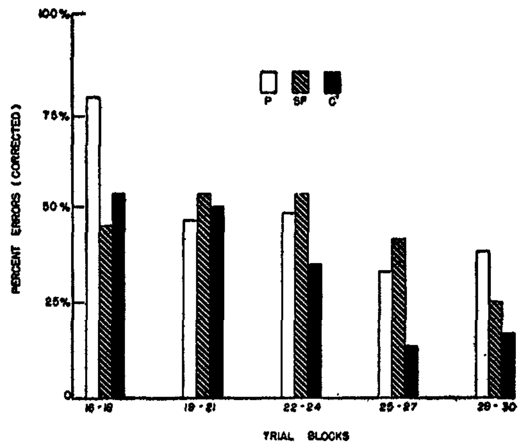

Fig. 3. Corrected Mean Percentage Errors for Experimental Groups in Part II

Results

Measurement of Coding Noise

It would be most appropriate at this point to measure the coding noise in this experiment in terms of the conditional entropies of information theory (6). However, to apply

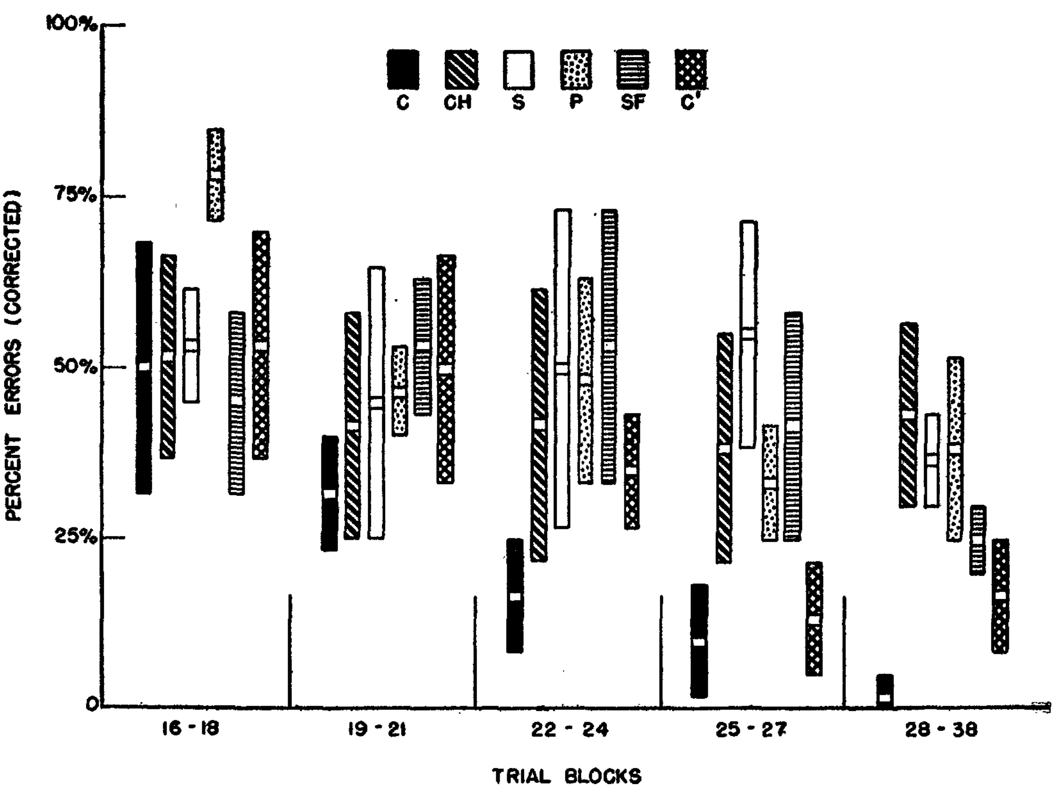

Fig. 4. Confidence Limits at $p = .05$ for the Mean Percentage Errors, All Experimental Groups

information theory measures to the coding noise occurring in this experiment requires knowledge of several facts which are unavailable. We are unable to specify the set of symbols an *S* would use to describe a given color marble, and, more important, we have no way of observing the transformations from the given set of marbles to the messages describing them, nor the corresponding mental transformations on the part of the receiver from the symbols of the received messages to the possible set of six marbles. Consequently, although we may discuss the noise in this experiment in terms of the concepts of information theory in a qualitative manner, we are unable to arrive at the usual quantitative measures. Nevertheless, a numerical measure is needed, and it may be achieved in an approximate manner by considering more closely the characteristics of the noise occurring in the experiment.

Since noise is fundamentally a question of uncertainty, any single valued measure of the amount of uncertainty in an experiment can be expected to be monotonically related to the noise. In this experiment, the uncertainty arose largely from different *S*s' applying the same name to different marbles, with the result that comparing the marble color names used by each *S* to describe the marbles he had at the trial led to several possible answers or, in some cases, to a single incorrect answer. Specifically, during the first 15 trials the groups generally learned to refer to each marble by a single color name, such as "red," "black," etc. After the sixteenth trial, even though the marbles used were mottled and streaked, often with more than one color, or with shades of one color, this behavior persisted. The *S*s usually attempted to use one-word color names, such as "amber," "aqua," or in some cases compound words such as "light-green" or "blue-green," to describe the marbles. These considerations led to the following procedure which was used to calculate this uncertainty (referred to as the "ambiguity," or marbles per name, and denoted by A). For each trial, the message cards sent by any *S* were examined, and in all cases in which a definite assignment of names to marbles could be made, on the basis of *E*'s knowledge of the marbles in each man's box, this information was tabulated.

From these results for all five *S*s for that trial, lists of names which had been used to describe each marble were compiled with the frequency of occurrence of each, and the weighted average of the number of marbles referred to by a name was calculated. This procedure was followed for all the groups run, for trials 16 to 30. These values were corrected as follows: If a given name was used to describe two different marbles on trials $i-1$ and $i+1$, but specific evidence for this confusion could not be found during trial i, it was assumed to be present on the strength of its occurrence before and after trial i. From these corrected values of A, an average value was computed for each network during each block of three trials. These results are presented in Fig. 5.

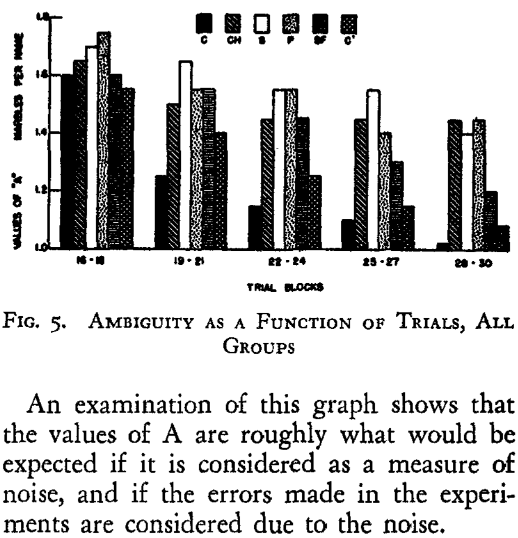

FIG. 5. AMBIGUITY AS A FUNCTION OF TRIALS, ALL GROUPS

An examination of this graph shows that the values of A are roughly what would be expected if it is considered as a measure of noise, and if the errors made in the experiments are considered due to the noise.

It is possible to relate A to the semantic noise level calculated on the basis of source and receiver entropies if a few very questionable assumptions are made. If the source in all cases uses a code of six symbols to refer to the six marbles, and each of these six symbols occurs equiprobably, then the entropy of the source is

$$H(x) = -\sum_i p(i) \log_2 p(i) = 2.59 \text{ bits per symbol.}$$

Since the observed average value for A lies between 1.0 and 2.0, let us assume that all the ambiguity for each received semantic symbol lies in a choice between two possible referents. If a symbol α is received, and A is given, the probability that the referent α will be chosen is

$$P_\alpha(\alpha) = 1/2\,(3-\text{A}),$$

which is a linear function such that when $A=1.0$, $P_a(\alpha)=1.0$ and when $A=2.0$, $P_a(\alpha)=1/2$. If β is the other possible referent, then

$$P_a(\beta)=1-P_a(\alpha).$$

The average conditional entropy of the receiver, when the source message is known, is defined as

$$H_x(y)=-\sum_{i,j} P(i,j)\ log_2 P_i(j)$$
$$=-\sum_i P(i) \sum_j P_i(j)\ log_2 P_i(j)$$

where $P_i(j)$ is the probability of picking marble j when symbol i is sent. We have assumed

$$P(i)=1/6$$
$$P_i(i)=1/2(3-A)$$
$$P_i(j_1)=1-P_i(i)=1/2(A-1)$$
$$P_i(j)=0 \text{ for } j\neq i,\ j\neq j_1.$$

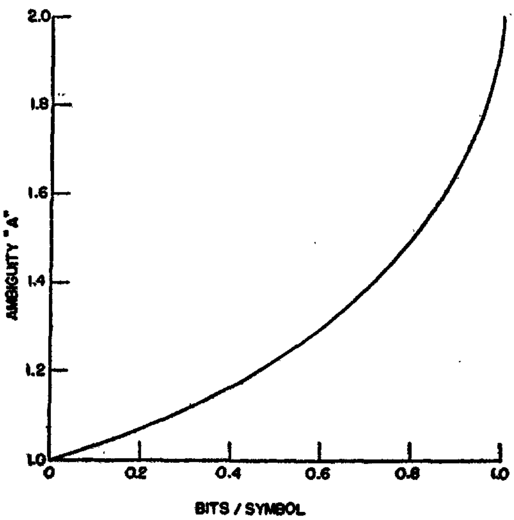

FIG. 6. AMBIGUITY AS A FUNCTION OF ESTIMATED CONDITIONAL RECEIVER ENTROPY

Using these values in the expression above, we get

$$H_x(y)=1-1/2\ [(3-A)\ log_2\ (3-A) + (A-1)\ log_2\ (A-1)].$$

$H_x(y)$ is plotted as a function of A from 1.0 to 2.0 in Fig. 6. The relationship between these two measures is not linear, and would probably be even less linear if calculated on the basis of a knowledge of the actual number of referents among which each choice was made.

The calculation given above, of course, does not constitute a *measured* value of $H_x(y)$—the values of A are measured, but $H_x(y)$ was calculated from A only after making several assumptions about the coding processes which cannot be verified from the data. Future experiments will be designed to make direct measurement of $H(x)$, $H_x(y)$, and $H(y)$ possible.

Errors as a Function of Coding Noise

It is possible by a few calculations based on some simplifying assumptions to confirm the hypothesis that the errors in this experiment are directly related to the noise level as measured by A. Consider an *S* who has received sufficient messages to reach a decision on an answer. Uncertainty in the coding process renders this decision ambiguous. This ambiquity will be idealized by assuming that the final choice is between two marbles, and that

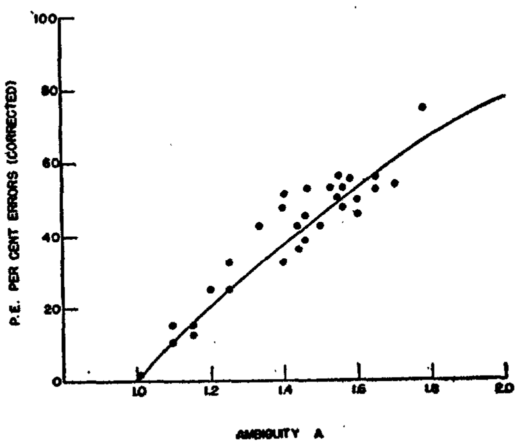

FIG. 7. PREDICTED PERCENTAGE ERRORS AS A FUNCTION OF AMBIGUITY

the probability of picking the correct one is $P=1/2(3-A)$, as above. This answer must then be sent to the other *S*s, and we make the assumption that any one of these others has a probability *P* of picking the correct marble when sent a description of it. The chance of a group member's dropping the correct marble is then P^2, except for the *S* who made the original choice and sent out the answer, who has a probability *P* of dropping the right marble. The average predicted percentage of error is then

$$E'=\frac{100}{5}[4(1-P^2)+(1-P)]$$
$$=10[13A-2A^2-11].$$

The values of the predicted error are plotted in Fig. 7, as the heavy line, and the observed percentages of error for each trial block are plotted as points, against the measured ambiguity for that trial block. These values fall along the predicted curve about as closely as we could ask, considering the variability of the data and the fact that the predicted line is not a fitted curve. The agreement of the observed points with this predicted line gives confirmation of the hypothesis that the errors are the result of semantic noise in the coding process.

Redundancy as a Mechanism for Reducing Errors

Since certain networks manage to achieve a reduction in the ambiquity, and hence in their error level, one is led to inquire about the mechanism of this effect. This problem may also be approached by an application of the concepts of information theory, extended to fit this case.

In the conventional case of signals transmitted along a channel, accurate transmission in the presence of noise is achieved at the expense of transmission rate by the introduction of redundancy. Since the noise here is semantic noise, we shall have to look for semantic redundancy, i.e., duplications in the coding scheme. In our case, these duplications, if they exist, will take the form of synonyms, or alternate descriptions of a given marble. We shall show that these duplications do exist and that they are used to overcome the noise.

Since the noise present in this experiment is semantic noise, and is measured by ambiguity A, the effect of the use of redundancy to overcome the noise and insure accurate transmission of the message will be to reduce or eliminate the apparent ambiguity or uncertainty present, and this will be reflected in a decrease in the measured value of A. In a sense, this case is not an exact parallel to the usual case of channel noise, since with channel noise the introduction of redundancy in the coding does not remove the noise, but merely removes the errors caused by the noise. Hence, in the channel-noise case, the redundancy must be maintained at a high level in order to insure accuracy. This is not the case with semantic or coding noise, for once the uncertainty in the coding operations has been eliminated, the redundancy may then be reduced without impairing the accuracy of the transmissions. However, it may also be thought of as having the constant character of channel noise by considering the effect of memory. Once the redundant coding has been used, and the errors reduced thereby, we may assume that the receiver remembers the synonyms used for a given symbol in the redundant code, and that in future messages these synonyms or alternate codes are understood even though not physically present. If the effect of this understood or remembered redundancy is assumed, we may describe the system as one with constant noise but with the effect of the noise overcome by the redundant coding, just as in the channel-noise case. In this description the redundancy is in two parts and that part attributable to memory does not appear in the transmitted message. Therefore, the total redundancy can remain high while the external redundancy drops. A decision on whether this is in fact the case will have to be the subject of a separate experiment, because we have data only on the external redundancy.

To detect semantic redundancy, we use a method analogous to that previously used to calculate ambiguity. In any one group, at any one trial, six names are sufficient to identify the six marbles. By tabulating from the message cards the names used by the group to describe a given marble, we obtained a record of synonyms or alternate codings used in each trial by each group. This tabulation was corrected for the ambiguity of some of the names used, on the basis that a synonym which was also applied to two other marbles should not be counted as a separate synonym for each, so the tabulation was weighted according to the ambiguity of each term. The table was also corrected for missing data—i.e., for a synonym which was used before and after a given trial, but evidence for the use of which could not be found with certainty during the trial. From these tabulations the average number of names used by each group during each trial was calculated, and from these figures the average number of extra names—that is, the number of names used beyond the necessary six—was calculated. These values were then averaged over all the groups run on a given network, and over blocks of three trials apiece, as previously.

The average number of extra names used is called the redundancy R, and is tabulated in Fig. 8 for the different networks, by blocks of three trials. Figure 8 supports the hypothesis that redundancy is used to overcome the errors due to semantic noise. Comparing Fig. 8 with Figs. 2 and 3, we see that those networks which reduced their error count show a rise in the redundancy subsequent to trial 16, and the reduction in errors always comes after this rise.

This dependence may be emphasized by a simple plot. Define two parameters, α and β, for each network at each trial block, as follows:

If this or a previous trial block had $R \geqq 3.6$, let $\alpha = 1$.
If all previous trial blocks had $R < 3.3$, let $\alpha = 0$.
If $A \leqq 1.4$ for this trial block, let $\beta = 1$.
If $A > 1.4$ for this trial block, let $\beta = 0$.

Define $\mu = \alpha + \beta$. Then, if a group has reduced its ambiguity by increasing R, μ should be 2. If the group has neither reduced A nor increased R, μ should be 0. An intermediate value ($\mu = 1$) will show contradictory behavior—a drop in A without a previous increase in R, or a rise in R which does not reduce A. The values of μ are plotted in Fig. 9 for each network and each trial block. The lack of values of $\mu = 1$ on this plot, and the points at which the values of μ jump to 2 indicate clearly that ambiguity is reduced in these groups by increasing the redundancy.

By the introduction of a measurement of redundancy, appropriately defined, we have thus been able to demonstrate the mechanism used by these groups to reduce their errors.

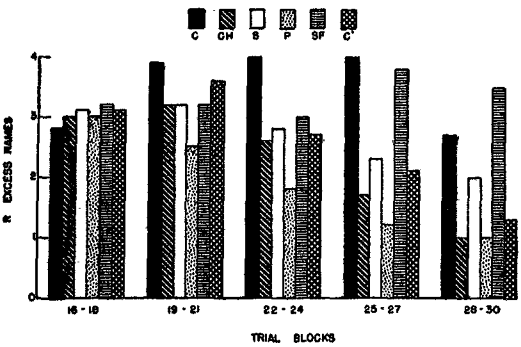

FIG. 8. REDUNDANCY AS A FUNCTION OF TRIALS, ALL GROUPS

The Effect of the Communication Network

It has been implicit in the discussion that the amount of redundancy, and hence the error reduction, is a function of the network. For example, the circle successfully achieved error reduction, whereas the chain and the star did not. Several conjectures arose from the results of Part I, and Part II was carried out as a preliminary attempt to verify these conjectures.

With regard to error reduction, two aspects of the communication network seem important: (*a*) that there are sufficient interconnections for each *S* to realize that the group makes errors, and (*b*) that there are sufficient interconnections (possibly different from those of *a*) for the *S*s to be able to correct the errors, once they are aware of them. The type of connections necessary for (*a*) are those that will allow, with fairly high frequency, an *S* to receive what purports to be the same information via two or more routes. If there is noise, then it is unlikely that a piece of information will travel two different routes to a given person and arrive there as the same symbol. Thus, in this class of network, one or more *S*s are very liable to receive two different symbols which purport to refer to the same marble, and this will certainly suggest to him that an error is being made. He may not be able to do anything about it during that trial, but in succeeding trials he can attempt to find ways to avoid further errors. Observe that the chain and the star are not in this class, since it is comparatively difficult for *S*s in these networks to know that errors

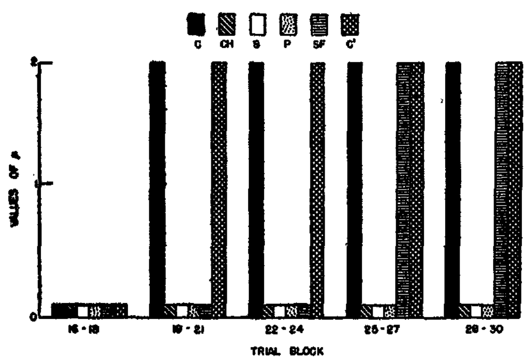

FIG. 9. PERFORMANCE PARAMETER μ AS A FUNCTION OF TRIALS, ALL GROUPS

are being made. On the other hand, the circle is in this class.

With regard to the second condition, it is in principle sufficient that each person be connected in the network to each other person; however, it is subjectively obviously that error correction between a and b is difficult if every message from a to b must pass through a third person c. This causes difficulty in asking questions of the sort, "What do you mean by aqua?", which are vital to the creation of redundancy. Thus we might suppose that symmetric communication channels are necessary to error reduction, assuming the existence of errors is known.

TABLE 1

QUALITATIVE SUMMARY OF NETWORK FACTORS AND ERROR REDUCTION

NETWORK	ERROR FEEDBACK	HIGHLY CENTRAL POSITION	SYMMETRIC CHANNELS	ERROR PERFORMANCE
Circle	Yes	No	All	Learns fast. Good error reduction.
Star	No	Yes	All	No learning. No error reduction.
Chain	Slight	Yes	All	No learning. No error reduction.
Pinwheel	Yes	No	None	Some initial learning. Poor error reduction.
Star with Feedback	Yes	Yes	All	Slow learning. Fair error reduction.

So, we consider two properties, the possibility of error feedback and symmetric channels. The circle has both, and the star only the latter. This suggested that we run a case having only the former, and pinwheel (P) was elected for this purpose. It has no symmetric channels, and, clearly, it has as much possibility of error feedback as the circle. Since we were not sure that these two factors take everything into account, we elected to run more star cases, but, in contrast to Part I, to give the *S*s the following information at the end of each trial: the number of errors made, and the number of different marbles dropped in error. These are the SF (star with feedback) groups of Part II. It is clear that this is at least as much information as the *S*s in circle groups would have, so that any difference between the circle and SF groups favoring the former is conservative. Finally, we ran the circle groups (C′) again in order to have a control group, since the *S*s in Part I were M.I.T. students and in Part II military personnel.

The experimental results have been presented in Figs. 3 and 4. First, let us look at the control group. The behavior is qualitatively the same as the Part I circles, simply displaced to the right. Presumably this results from differences between the two classes of *S*s in intelligence or motivation or both. We see that the pinwheel groups begin with a very large error count and reduce it to a level indistinguishable from the star and chain groups. This certainly suggests error knowledge, as expected, and some error correction through the three-step feedback loops, but by no means the degree of error correction achieved in the circle groups. Finally the SF groups do begin error correction in a fashion not unlike the circle, but markedly displaced to the right. As we pointed out, this is a conservative difference, so we cannot claim the performance of the SF groups is the same as that of the circle.

It must be concluded that though error feedback and symmetric channels are necessary to good error reduction, there is at least one other factor of importance. Observe that in the star the entire process of noise reduction must be located at the central position. The other *S*s contribute to this process, but only in a passive way, e.g., answering questions from the center man. In the case of both the circle and pinwheel it is possible for each *S* to participate actively in the noise reduction process, and that process may be carried out in comparatively small steps and either the entire group establish a common code, or each pair or trio of persons arrive at its own private code. If this conjecture is valid, we are simply observing a time delay due to job complexity at the center position of the star, with the same two basic mechanisms operating as we pointed out earlier.

We may summarize our knowledge as Table 1.

In conclusion, it should be pointed out that though we feel these factors to be important they are not stated with the precision we would like nor are they established with the definiteness we would like. It must be kept in mind that these experiments did not employ enough groups in each category to be conclusive. In addition, since the original experimental design was for quite another purpose, estimates and approximations were necessary in the analysis. As this seems an important area, we propose to redesign the experiment and carry out a more complete analysis in the near future.

SUMMARY AND CONCLUSIONS

An experiment involving task-oriented groups of five *S*s, requiring for problem solution the use of descriptions of similarly colored marbles, was analyzed in terms of the concepts of information theory. The errors made by the group were shown to be well predicted by a measure of the semantic noise in the coding-decoding process, and the use of redundant coding to reduce the number of errors was demonstrated. Differences in behavior of groups using different communication networks was examined, and several properties of the communication network which play important roles in determining this behavior were discussed. The application of information theory to this type of group experiment, and its extension to the problem of communication which is semantically noisy, was shown to be a valid and useful method for analyzing such experiments.

REFERENCES

1. BAVELAS, A. Communication processes in task-oriented groups. *J. acoust. Soc. Amer.*, 1950, **22**, 725–730.
2. BAVELAS, A., & BARRETT, D. Experiments in communication and organization. *Personnel*, 1951, **29**, 2–7.
3. CHRISTIE, L. S., LUCE, R. D., & MACY, J. *Communication and learning in task-oriented groups.* Cambridge, Mass.: Technical Report No. 231, Research Laboratory of Electronics, M.I.T., 1952; and Santa Monica, Calif.: Research Memorandum, the RAND Corp., 1952.
4. HEISE, G. A., & MILLER, G. A. Problem solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, **46**, 327–335.
5. LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38–50.
6. SHANNON, C. E. A mathematical theory of communication. *Bell System Tech. J.*, 1948, **27**, 380–423, 623–656.

Received June 23, 1952.