WADC TECHNICAL REPORT 54-362

THE INFLUENCE OF CATEGORICAL CUING ON THE IDENTIFICATION OF VISUALLY DISTORTED WORDS SELECTED FROM A LOGICALLY ORGANIZED POPULATION

The Seventh of a Series of Reports on
"SET" As a Determiner of Perceptual Responses

RICHARD H. HENNEMAN
L. STARLING REID
EUGENE R. LONG

UNIVERSITY OF VIRGINIA

APRIL 1955

WRIGHT AIR DEVELOPMENT CENTER

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Qualified requesters may obtain copies of this report from the ASTIA Document Service Center, Knott Building, Dayton 2, Ohio.

- - - - -

Copies of WADC Technical Reports and Technical Notes should not be returned to the Wright Air Development Center unless return is required by security considerations, contractual obligations, or notice on a specific document.

WADC TECHNICAL REPORT 54-362

THE INFLUENCE OF CATEGORICAL CUING ON THE IDENTIFICATION OF VISUALLY DISTORTED WORDS SELECTED FROM A LOGICALLY ORGANIZED POPULATION

The Seventh of a Series of Reports on
"SET" As a Determiner of Perceptual Responses

RICHARD H. HENNEMAN
L. STARLING REID
EUGENE R. LONG

UNIVERSITY OF VIRGINIA

APRIL 1955

AERO MEDICAL LABORATORY
CONTRACT No. W 33(038)-ac-21269
PROJECT No. 7192

WRIGHT AIR DEVELOPMENT CENTER
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

Carpenter Litho & Prtg. Co., Springfield,O.
100 - 27 February 1956

FOREWORD

This report is the seventh of a series relating to laboratory experiments designed to explore the operation of perceptual "set" and its application to human operator efficiency in the process of the intake of information. The investigation covered in the present report was conducted by Dr. R. H. Henneman, Dr. L. S. Reid, and Dr. E. R. Long. These studies were carried on at the University of Virginia under Project No. 7192, Task No. 71603, "Visual Message Presentation." The contract was administered by the Psychology Branch of the Aero Medical Laboratory, Directorate of Research, Wright Air Development Center, with James E. Smithson acting as Project Engineer.

ABSTRACT

In air communication the identification of ambiguous message signals is substantially aided by operator familiarity with the message population and by the high degree of contextual redundancy present. Previous laboratory research on perceptual setting as an aid to the identification of ambiguous messages has ignored the factors of familiarity and contextual organization. The present experiment constituted an attempt to learn whether further response restriction in the form of setting cues would facilitate the recognition of visually distorted words drawn from a logically organized population, when subjects were familiar with the word categories and the specific words.

Experimental variables were (1) degree of categorical restriction (in cuing), (2) type of familiarization (categories only or categories plus specific words), and (3) temporal position of the setting cues (before or after stimulus presentation). Principal findings were: (1) Increases in categorical restriction led to improved identification. (2) Familiarity with specific words, apart from a knowledge of the word categories, significantly improved identification. (3) Temporal position of cuing was not a significant factor. These results imply the probable helpfulness of setting cues for the identification of ambiguous messages in operational situations where familiarity and context have already markedly restricted the operator's responses.

PUBLICATION REVIEW

This report has been reviewed and is approved.

FOR THE COMMANDER:

JACK BOLLERUD
Colonel, USAF (MC)
Chief, Aero Medical Laboratory
Directorate of Research

TABLE OF CONTENTS

INTRODUCTION

In an earlier analysis of the communication process (4) it was suggested that the complex perceptual behavior involved in message reception might be considered as composed of chains or sequences of discriminatory response classes, such as stimulus detection, stimulus location, and stimulus identification. Frequently message stimuli, because of such factors as impoverishment, distortion, or field "clutter", are capable of occasioning more than one response in a particular response class, i.e., they give rise to a situation of response ambiguity or uncertainty. It was further suggested that setting cues (defined as response-restricting information) should lead to more accurate message reception by increasing the probability of certain responses and decreasing the probability of other competing responses.

A series of exploratory investigations reported in earlier technical reports (2, 3, 4, 5, 6, 7) has generally confirmed this theoretical analysis. Response-restricting cues have been shown to facilitate the location and identification of geometrical figures and letters of the alphabet. More recent experiments have sought to extend the generalization of the earlier findings by introducing additional experimental variables applying to actual operational situations. Thus, in the most recent experiment (3), words replaced the earlier figures as stimulus materials, voice presentation was included in addition to visual input, and recall measures of message intelligibility were employed as well as recognition.

It was the purpose of the present investigation to extend still further the generalization of the previously indicated helpfulness of perceptual set in situations calling for the identification of ambiguous message signals. It was held advisable to obtain information concerning the effectiveness of "setting" the operator under conditions more nearly approximating those encountered in air operations, conditions obviously differing in some instances from those of the laboratory situation. For example, message analysis research in both civilian and military air traffic control operations has revealed the highly structured nature of air communication. The messages are clearly categorized, and visually follow one another in definite sequences. Moreover, the operators who send and receive these messages are not only highly familiar with the general message population but also enjoy the advantages of contextual setting.

Indeed it seems likely that this message familiarity and contextual setting contribute principally to the success with which the veteran operator is able to interpret the garbled sound signals coming through his earphones. For example, whereas Shannon has estimated that the English language in general is about 75% redundant (9), Frick and Sumby (1), point out that such air language statements as "Air Force 5264, ready number one in take-off position; over," when further constrained by the usual situational

and procedural contexts, has an estimated redundancy of approximately 96%. Laboratory evidence for the importance of context in message intelligibility comes from Miller, Heise, and Lichten (8), who have demonstrated that the context produced by (1) restricting the size of message vocabulary, (2) combining words into sentences, or (3) repeating messages, greatly increases the redundancy of the message and thus increases the probability of correct identification for a given signal-to-noise ratio.

These considerations led the present investigators to speculate as to whether or not the findings of the earlier experiments at Virginia would generalize to the highly structured and familiar messages exchanged during communication in air operations. More specifically, it was questioned whether further attempts at response restriction by means of setting cues would be effective in a perceptual situation in which responses were already markedly restricted by reasons of message familiarity and context. Accordingly, another laboratory investigation was designed to answer the following questions.

(1) Can the probability of identification of visually distorted words be increased by the progressive restriction of the specific categories in which they occur within a logically organized word population?

(2) If progressive categorical restriction does increase the probability of correct identification, is this increase contingent upon a particular temporal pattern of cuing?

(3) Is familiarity with the specific words, apart from knowledge of the word categories, a factor in the effectiveness of this "setting"?

(4) Is there a critical point in the degree of response restriction where further restriction fails to aid word identification?

In seeking answers to these questions the following conceptual variables were manipulated: (1) degree of categorical restriction, (2)type of previous stimulus familiarization, and (3) temporal position (with respect to stimulus presentation) of the cuing. Operationally these variables were employed in conjunction with distorted words, presented visually, which subjects were required to identify. Cuing, designed to reduce identification response uncertainty, was carried out by presenting the subjects with the names of categories which contained the stimulus words. Degree of setting was varied by increasing the number of category or descriptive names---the more names, the greater the categorical restriction, that is the smaller the sub-category hence the smaller the number of stimulus words included. The presentation of these identification cues was made before, after, or both before and after stimulus presentation. Stimulus familiarization was varied by giving two types of training prior to attempted identification of the stimuli. One group was allowed only to study the organizational structure of the categories; the other group was allowed in addition to study the specific stimulus words themselves.

PROCEDURE

Experimental Design

The variables---degree of categorical restriction, type of stimulus familiarization, and temporal pattern of cuing---were organized into a 5 x 2 x 3 factorial design. Table 1 depicts all combinations of the three variables.

Subjects

Ninety subjects (male college students) were randomly assigned to the 30 cells (3 per cell), representing all combinations of the experimental conditions, as shown in Table 1.

Preparation of Stimulus Material

The stimulus population was composed of 48 words all related to sports. The stimulus words were prepared for presentation by first typing them on individual cards. These words were made of equal length by typing x's before the first letter and after the last letter of each words, as shown in the following examples: X D O D G E R S X; X X X C O B B X X. The resulting word patterns were then photographed with camera misfocused. This procedure resulted in the production of 35 mm transparencies which, when projected out of focus, were correctly identified only 2% of the time under conditions of no setting, i.e., without category cues or previous study of the words.

Manipulation of Stimulus Familiarization

The 48 stimulus words were selected so that they were capable of being subdivided into successively smaller categories. Thus although all the words were related to sports, one-half were related to baseball and the other to football. Each of these categories (each containing 24 words) was further subdivided into the two subcategories of names and terms. These were then further broken down into teams, players, positions, and plays, and so on. Figures 1 and 2 indicate all of the subcategories and the specific stimulus words which fell in each category.

Prior to the subjects' attempted identification of the words, two types of familiarization were provided. One group of subjects was given a sheet on which both the organization of the categories and the stimulus words were printed. The second group was given sheets which contained only the organization of the categories, without the words. Both groups

TABLE 1

Design of Experiment

Degree of Setting

Type of Familiarization	I		II		III		IV		V	
	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only
Pre-Stimulus	3	3	3	3	3	3	3	3	3	3
Post-Stimulus	3	3	3	3	3	3	3	3	3	3
Pre-Plus Post-Stimulus	3	3	3	3	3	3	3	3	3	3

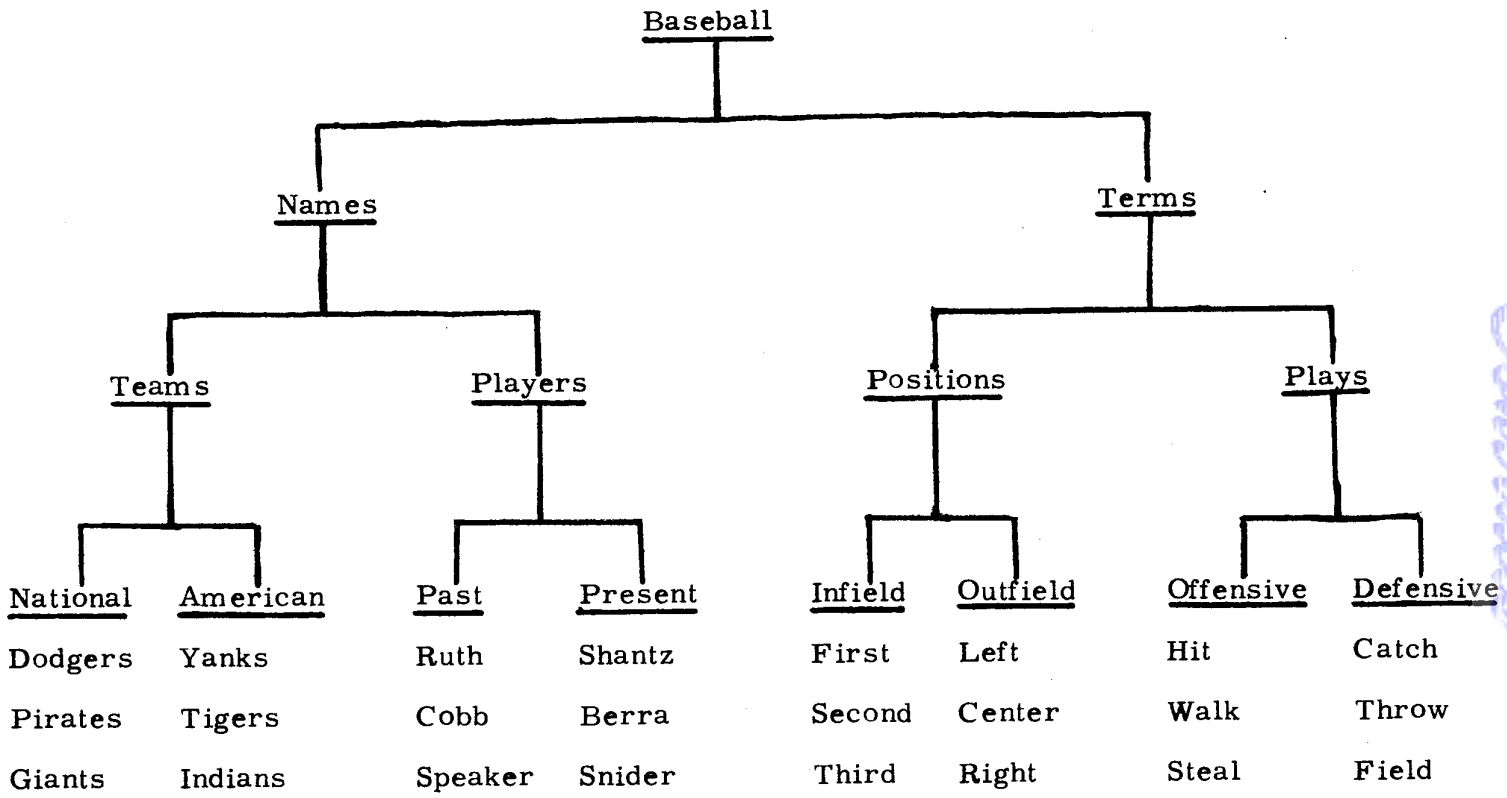

Figure 1. Organization of the First Principal Category Under the General Class, Sports.

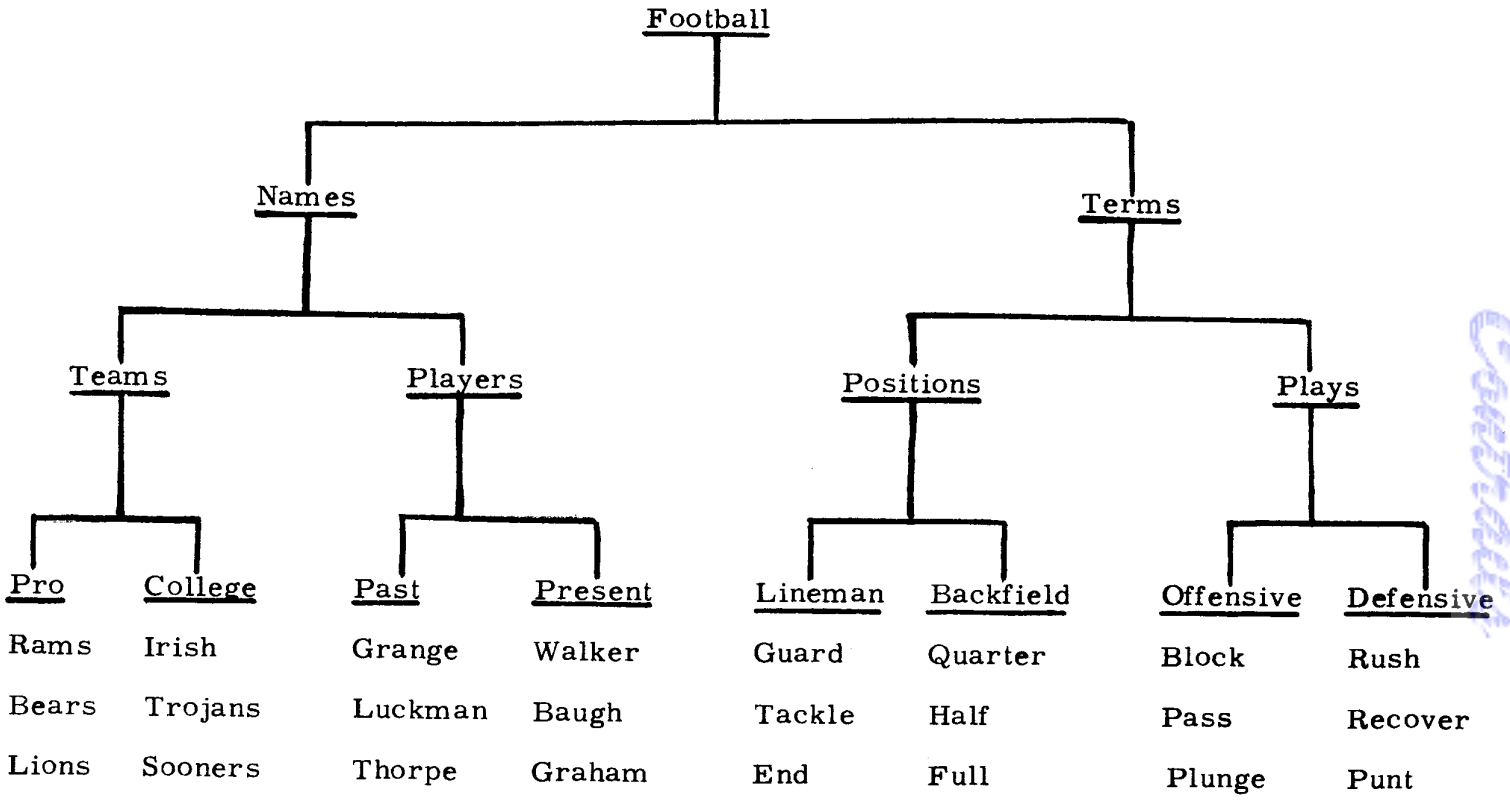

Figure 2. Organization of the Second Principal Category Under the General Class, Sports.

were given a five-minute period during which they studied their respective sheets. At the end of that time they were given a written recall test. Following this test they were given a second five-minute study period, at the end of which time stimulus presentation was begun.

Manipulation of Setting Cues

During the course of stimulus presentation, degree of setting was varied by giving the subject varying numbers of category terms, ranging from one to five. Thus, those subjects receiving the lowest degree of setting saw only one word, SPORTS, printed on their cuing sheets. Those receiving the next highest degree of setting saw two words (SPORTS - BASEBALL or SPORTS - FOOTBALL), depending upon which stimulus word was presented. Those subjects having the highest degree of setting saw five successive descriptive or categorizing words, thus, SPORTS - FOOTBALL - NAME - TEAM - PRO, the specific combination of cue words again being determined by the particular stimulus word being projected.

Experimental Procedure

The general experimental procedure (immediately following the familiarization period) was arranged so that each subject might receive one of the 15 combinations of degree of setting and temporal position of cuing. This was done in the following manner. During the first seven seconds of a particular trial, the subjects receiving pre-stimulus cuing were allowed to view an appropriate cuing sheet; those receiving no pre-stimulus cuing looked at a blank sheet of paper. At the end of this time those subjects receiving pre-stimulus cuing turned their cue sheets over. The stimulus was then presented for five seconds, and immediately following this those subjects receiving post-stimulus cuing exposed and examined their cuing sheets. Those receiving no post-stimulus cuing sat quietly for the seven seconds allowed for post-stimulus cuing. At the end of this time, those being cued turned over their cuing sheets, and all subjects were then required to write down what they thought the stimulus word had been. Ten seconds were allowed for this. This procedure was repeated 48 times, once for each stimulus word, and required a total time of approximately 25 minutes.

RESULTS

The mean frequencies of correct identification and the combinations of conditions under which they were obtained are presented in Table 2. An analysis of variance (Table 3) performed on these data indicates that (1) increases in categorical restriction resulted in significant increases in

TABLE 2

Mean Number of Correct Identifications for Each Condition

Type of Familiarization	Degree of Setting I		II		III		IV		V	
	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only	Words Plus Org.	Org. Only
Pre-Stimulus	14.0	12.7	18.0	13.3	21.7	15.7	24.3	31.7	42.0	17.0
Post-Stimulus	19.7	15.0	16.3	13.3	16.7	16.3	30.0	26.0	31.0	12.3
Pre-Plus Post-Stimulus	19.0	18.7	16.7	12.0	24.0	17.0	37.0	24.3	32.0	25.0

TABLE 3

Analysis of Variance of Correct Identifications

Source of Variation	Sum of Squares	df	Mean Square	F	p
A. Degree of Setting	2,807.2	4	701.8	13.3	.01
B. Temporal Pattern of Setting	126.3	2	63.2	1.2	
C. Type of Familiarization	846.4	1	846.4	16.0	.01
A x B	269.8	8	33.7		
A x C	665.7	4	166.4	3.2	.05
B x C	0.6	2	0.3		
A x B x C	609.3	8	56.2	1.1	
Error	3,170.0	60	52.8		
Total	8,495.3	89			

frequency of correct identification; (2) previous familiarization entailing both the stimulus words and the categorical organization produced significantly more correct identification than did familiarization with the categorical organization alone; and (3) temporal pattern of cuing did not significantly influence identification.

Because of a significant interaction between degree of setting and type of familiarization, separate analyses of variance were carried out (Table 4). These analyses indicate that under either type of familiarization, frequency of correct identification was not significantly influenced by temporal pattern of cuing. It was significantly influenced in both instances, however, by degree of setting. The important thing to note is that in the case of that group which had been previously familiarized with both words and organization, almost all of the sum of squares contingent upon increases in categorical restriction was contributed by one degree of freedom --- that associated with the linear effect. In the case of that group previously

TABLE 4

Analyses of Variance of Correct Identification Scores With Type of Previous Familiarization Held Constant

I. Familiarization Entailing Both Stimulus Words and Categorical Organization

Source of Variation	Sum of Squares	df	Mean Square	F	p
A. Degree of Setting	2,370.0	4	592.5	10.3	.001
Linear Regression	2,102.5	1	2,102.5	36.7	.001
Residual	267.5	3	89.1'	1.4	
B. Temporal Pattern of Setting	68.0	2	34.0		
A x B	542.0	8	67.8	1.2	
Error	1,720.0	30	57.3		
Total	4,700.0	44			

II. Familiarization Entailing Only Categorical Organization

Source of Variation	Sum of Squares	df	Mean Square	F	p
A. Degree of Setting	1,103.0	4	275.7	5.7	.005
Linear Regression	360.0	1	360.0	7.4	.025
Residual	743.0	3	247.6	5.1	.01
Quadratic Regression	19.8	1	19.8		
Residual	723.2	2	361.6	7.5	.005
Cubic Regression	618.8	1	618.8	12.8	.005
Residual	104.4	1	104.4	2.1	
B. Temporal Pattern of Setting	59.0	2	29.5		
A x B	337.0	8	42.1		
Error	1,452.0	30	48.4		
Total	2,951.0	44			

familiarized with only the organization, however, two degrees of freedom made the major contributions to the sum of squares contingent upon increases in categorical restriction. These were the degrees of freedom associated with the linear and cubic effects. An inspectional analysis of the group means reaffirms this, suggesting that even though there was an average increase in correct identification with increased setting, the increase was not a regular one, and because of the additional cubic effect a point was reached beyond which increased setting resulted in actual decrements in performance.

DISCUSSION

The questions previously posed in the introduction may now be answered in the light of the obtained results.

1. The probability of correct identification of visually distorted words can be increased by progressive restriction of the specific categories in which they occur within a logically organized word population. This was clearly indicated by the fact that increases in categorical restriction were generally accompanied by increases in frequency of correct recognition.

2. The facilitating effect of categorical restriction was independent from temporal pattern of cuing. Thus, pre-, post-, and both pre- plus post-cuing neither significantly influenced the data independently nor significantly interacted with degree of setting.

3. Familiarity with the specific stimulus words, apart from a knowledge of the word categories, is a factor in the effectiveness of the setting. This is indicated by two findings. First, that group which was acquainted with both the words and organization was significantly superior to the group which previously saw only the organization. Second, the group which previously saw both words and organization showed a general linear increase in frequency of correct identification as the degree of categorical restriction was increased. The group familiar only with the organization also showed a linear increase in frequency of correct identification with progressive categorical restriction, but to a lesser degree. Thus performance at the highest level of setting was significantly superior to that at the lowest for both groups. On the other hand, the performance of the latter familiarization group (i. e., acquainted with organization only) at the highest level of setting was inferior to that at the level of setting immediately below the maximum. This result appears to be due to that degree of freedom associated with the cubic regression effect, previously indicated in the analysis of results. This finding suggests that the facilitating effect resulting from progressive categorical restriction built up to a maximum and then decreased with further restriction.

4. No final answer can be given to Question 4 at this point. The several interpretations of the question call rather for the consideration of a number of alternatives. The data just discussed in Question 3 suggest that there is a critical point where further restriction not only fails to aid identification, but even inhibits it. Admittedly this result is not completely explicable. It would certainly seem reasonable to assume that the performance at maximum cuing restriction should be no worse than that at the level of cuing immediately below the maximum. Therefore, possible failures to meet assumptions of random sampling, etc. must be seriously considered. On the other hand, it may be of some significance that this detrimental effect occurred in the case of the group receiving only previous categorical organization familiarization, and not in the case of the group previously familiarized with both stimulus words and organization. It may well be that the type or degree of past familiarization fixes the practical limits beyond which increased setting is not only useless but also is deleterious. Because of the practical importance of this result and its implications, future research should be conducted in order to determine whether it is a reliable phenomenon. If it is found to be reliable, the conditions which underlie it (such as previous verbal habits, etc.) should be determined.

There is a second aspect to Question 4 which cannot be completely answered by the present data, but which should also be considered at this point in the light of its practical significance. The point was made in the introductory section that in the practical situation there was (1) so much categorical and sequential structure in the message language, (2) so many non-message (contextual) stimuli which reduce uncertainty, and (3) so much familiarity with the message population, that attempts at further restriction by the use of setting cues would be useless. In other words, so much non-message response restriction already existed that any further reduction in uncertainty would have to come through an improvement in the message stimuli themselves, rather than through any form of further cuing of the operator. This line of reasoning leads to the conclusion that there must be a critical point in response restriction mediated by setting. The present results in part dispute this conclusion. They indicate that even where the message stimuli are categorized and where the subjects have been familiarized with both the message structure and the specific population of messages, progressive response restriction by means of setting cues is still effective. As a matter of fact, the efficacy of progressive setting seems to be positively related to the type and degree of past familiarization (setting under the condition of familiarization with both words and organization being significantly superior to familiarization with organization only).

On the other hand, it might be argued that the degree of message structuring or the level of familiarization in the present study is far below that of the practical situation. For example, those subjects who received past familiarization entailing both words and organization but who were not differentially set, had a total correct identification score of only 158 out of a possible 432. The major point of this argument is that if these same subjects (because of greater familiarization with the message structure and

content) had been able to achieve correct identification scores of perhaps 390 to 400, then added cuing by categorical restriction or any other type of cuing would have been useless. Rather it would be the level of distortion of the message stimulus which was now the limiting factor, and it would only be through a reduction in stimulus distortion that any further enhancement of identification could be brought about. In answer to this it must be admitted that the level of distortion does set the limit to which setting may be practically employed. It does not necessarily follow, however, that the uncertainty reduction brought about by relatively high degrees of past familiarization or learning cannot always be further increased by setting cues. Further research entailing greater variation of both past familiarization and categorical restriction needs to be done in order to provide a more definitive answer to this important question.

SUMMARY AND CONCLUSIONS

Message analysis research has shown that in many air operations (e.g., traffic control) messages are clearly categorized and tend to fall into definite sequences. The operators who exchange these messages are not only familiar with the population of messages, but are usually aware of the specific category from which the next message is likely to be transmitted. The fact that in practical situations the operators are called upon to identify messages from familiar contexts raises important questions, hitherto not considered, as to the effectiveness of setting cues. It might be argued that in the process of receiving familiar, categorized messages further limitation of responses by cuing the operator would be of no advantage for message identification. Accordingly a laboratory investigation was designed to answer the following specific questions. (1) Can the probability of a given identification response be increased by the progressive restriction of the specific categories in which it occurs within a logically organized word population? (2) Is familiarity with the specific words, apart from knowledge of the word categories, a factor in the effectiveness of this setting? (3) If progressive categorical restriction does increase the probability of correct identification, is this increase contingent upon a particular temporal pattern of cuing? (4) Is there a critical point in the degree of response restriction where further limiting fails to aid word identification?

In answering these questions the following conceptual variables were manipulated: (1) degree of categorical restriction, (2) type of stimulus familiarization, and (3) temporal position (with respect to stimulus presentation) of the cuing. Operationally these variables were employed in conjunction with distorted words, presented visually, which subjects were required to identify. Cuing, designed to reduce identification response uncertainty, was provided by presenting the subjects with the names of categories containing the stimulus words. Degree of setting was varied by progressively reducing the size of the category, hence the number of stimulus words included. Presentation of these identification cues was made

before, after, or both before and after stimulus presentation. Stimulus familiarization was varied by giving two types of training prior to attempted identification of the stimuli, one limited to the organizational structure of the categories, the other including in addition the specific stimulus words.

Principal findings were the following:

(1) Increases in categorical restriction were generally accompanied by increases in frequency of correct word recognition.

(2) Familiarity with the specific stimulus words, apart from a knowledge of the word categories, is a factor in the effectiveness of the setting. The group familiar with both the words and organization was significantly superior to the group which previously saw only the organization. Both familiarization groups showed significant general linear increases in frequency of correct identification as the degree of categorical restriction was increased.

(3) The facilitating effect of categorical restriction was independent from temporal pattern of cuing. Thus, pre-, post-, and both pre- plus post-cuing neither significantly influenced the data independently nor significantly interacted with degree of setting.

(4) There is no clearcut answer to Question 4. The data for the group with only organization familiarity suggested that there was a critical point where further restriction not only failed to aid identification but even inhibited it. This result was not found for the other group (familiar with both words and organization). This difference would seem to indicate that the greater the amount of familiarization, the more evident the helpfulness of setting. However, this conclusion is tempered by the possibility that the degree of message familiarity achieved in the present experiment was well below that obtaining in practical operational situations. Only further research can provide a definitive answer to the probably complex relation existing between degree of message familiarity and categorical organization on the one hand, and effectiveness of setting on the other.

BIBLIOGRAPHY

1. Frick, F. C., and Sumby, W. H. "Control Tower Language." Journal of the Acoustical Society, Volume 24, 1952, pp. 595 – 596.

2. Long, E. R., and Garvey, W. D. "The Role of Setting Cues in Reducing the Simultaneous Location and Identification Ambiguity of Letter Patterns." The Fifth of a Series of Reports on "SET" as a Determiner of Perceptual Responses. WADC Technical Report No. 54 – 289, April 1955.

3. Long, E. R., and Garvey, W. D. "The Influence of Sensory Channel of Cuing on the Identification of Aurally Presented Distorted Words." The Sixth of a Series of Reports on "SET" as a Determiner of Perceptual Responses. WADC Technical Report No. 54 – 361, April 1955.

4. Long, E. R., Henneman, R. H., and Reid, L. S. "Theoretical Considerations and Exploratory Investigation of 'SET' as Response Restriction." The First of a Series of Reports on "SET" as a Determiner of Perceptual Responses. WADC Technical Report No. 53 – 311, December 1953.

5. Long, E. R., and Lee, W. A. "The Role of Spatial Cuing as a Response-Limiter for Location Responses." The Second of a Series of Reports on "SET" as a Determiner of Perceptual Responses. WADC Technical Report No. 53 – 312, December 1953.

6. Long, E. R., and Lee, W. A. "The Influence of Specific Stimulus Cuing on Location Responses." The Third of a Series of Reports on "SET" as a Determiner of Perceptual Responses. WADC Technical Report No. 53 – 314, December 1953.

7. Long, E. R., Reid, L. S., and Garvey, W. D. "The Role of Stimulus Ambiguity and Degree of Response Restriction in the Recognition of Distorted Letter Patterns." The Fourth of a Series of Reports on "SET" as a Determiner of Perceptual Responses. WADC Technical Report No. 54 – 147, May 1954.

8. Miller, G. A., Heise, G. A., and Lichten, W. "The Intelligibility of Speech as a Function of the Context of the Test Materials." Journal of Experimental Psychology, Volume 41, 1951, pp. 329 – 335.

9. Shannon, C. E. "Prediction and Entrophy of Printed English." Bell System Technical Journal, Volume 27, 1951, pp. 50 – 64.