code 441
NR201-272

LEVEL II

(12)
SC

AD A063192

DDC FILE COPY

(6) Multivariate Analysis of Evoked Potentials
and Semantic Meaning,

Final Report for the Period from
October 1, 1976 to July 31, 1978

(9) Final rept. 1 Oct 76-31 Jul 78,

Prepared for: Office of Naval Research
and
ARPA Human Resources Research

Sponsored by: Advanced Research Projects Agency
(15) Contract N00014-77-C-0037,
ARPA Order [redacted] 3294

Principal Investigator: Robert M. Chapman
(10)
Center for Visual Science
Psychology Department
The University of Rochester
Rochester, New York 14627

(11) 30 Sep 78 (12) 98p.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

Report Date: September 30, 1973

407 469

12.12.78

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
MULTIVARIATE ANALYSIS OF EVOKED POTENTIALS AND SEMANTIC MEANING

5. TYPE OF REPORT & PERIOD COVERED
Final Report
1 Oct 76 to 31 Jul 78

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Robert M. Chapman

8. CONTRACT OR GRANT NUMBER(s)
Contract
N00014-77-C-0037 new

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Center for Visual Science
University of Rochester
Rochester, New York 14627

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
ARPA Order No. 3294

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
800 N. Quincy Street
Arlington, Virginia 22217

12. REPORT DATE
30 Sept 78

13. NUMBER OF PAGES
33

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for Public Release, Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
Publications listed on reverse side.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Communication, Information Processing, Evoked Potentials, EEG, Brain Responses, Semantic, Memory, Neurolinguistics, Connotative, Number Concepts, Short-Term Memory, Storage Component, Psycholinguistics, Meaning, Biocybernetics

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
See other side.

78 11 27 052

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

18. (1) Chapman, R. M. On the road to specific information in evoked potentials. In Armington, J. C., Krauskopf, J. and Wooten, B. (Eds.), Visual Psychophysics: Its Physiological Basis. New York: Academic Press, 469-480. (2) Chapman, R. M. Connotative meaning and averaged evoked potentials. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior. New York: Plenum Press. (3) Chapman, R. M., McCrary, J. W. and Chapman, J. A. Short-term memory: The "storage" component of human brain responses predicts recall. Science, 1978 in press. (4) Other publications are anticipated.

20. The research provides confirmation that Evoked Potentials contain information about the semantic meaning of word stimuli used to obtain them and that combinations of Evoked Potential components can identify the unknown semantic character of stimuli. Previous results involving the use of relatively "pure" semantic stimuli have been successfully extended to words which are semantically more "complex."

The detection of numerical information by means of Evoked Potentials was explored, and tentatively confirmed, by analyzing the responses to visual number symbols and a 'think' cue (=). For both semantic and numerical meanings, the discriminant functions developed for a group of subjects worked equally well for each individual.

An Evoked Potential component with a post-stimulus peak about 250 msec. was found to be related to storage of information in short-term memory. In a behavioral experiment which probed short-term memory, recall was predicted by the magnitude of the Storage Component.

78 11 27 052

Multivariate Analysis of Evoked Potentials and Semantic Meaning

ABSTRACT

(I) The Representation of Combinations of Connotative Meaning in Brain Potentials

The effects of two kinds of experimental manipulation of semantic meaning were studied in Evoked Potentials (EPs), brain responses recorded from scalp monitors. Both kinds of semantic manipulation were based on Osgood's analyses which found three primary dimensions of connotative meaning: Evaluation, Potency, and Activity (E, P, and A). One kind of experimental variable was the semantic class of the stimulus word: High (E+,P+,A+), Neutral (E0,P0,A0), Low (E-,P-,A-). The other kind of experimental variable was the semantic dimension of the semantic scale (E, P, A) which the subject used to make semantic judgments about the stimulus words. These variables were experimentally combined so that for each trial the subject was using a designated semantic scale to judge a specified stimulus word while brain activity was recorded. Using multivariate procedures, both stimulus word class and scale dimension effects on the EPs, as well as their interaction, were analyzed.

Using EP measures, the 3 word classes used in this experiment were about as discriminable (pairwise) as the 6 word classes lying in other regions of semantic space which were previously reported. Thus, the generality of discriminating connotative meaning with EP measures has been confirmed with additional words belonging to different regions of Osgood's semantic space.

Common sets of classification functions were successful for the group of 13 individuals. This finding further supports the similarity of the EP effects in different individuals.

Simultaneous identification of word class and scale dimension was achieved at better than chance levels. Analyses indicated that these two kinds of semantic effects in EPs did not strongly interact and were largely independent. The semantic features of both words and tasks appear to be ascertainable either simultaneously or separately and appear to be relatively independently represented in the EP.

Separate analyses identified word class and scale dimension at better than chance levels. The same classification functions were successfully used for all subjects.

The evidence indicates that two kinds of semantic information are available in EPs: (1) processing of the semantic meaning in words, regardless of the semantic expectancies of the subject, and (2) semantic expectancies or judgment dimensions of the subject,

regardless of the semantic content of the words. The first kind of semantic information is more strongly represented in EPs than the second kind.

The results of these analyses of the data provide additional confirmation of the findings of previous phases: EPs contain information about the semantic meaning of word stimuli used to obtain them and that combinations of EP components show promise in identifying the unknown semantic character of stimuli which have evoked particular brain responses. Our previous results which involved the use of relatively "pure" semantic stimuli have been successfully extended to words which are semantically more "complex." Instead of being defined in terms of a single connotative dimension, the meanings represented combinations of the Evaluative, Potency, and Activity Dimensions. Word classes which were selected to be positive, negative, or neutral on all three dimensions simultaneously were reliably discriminated by EP analyses. We applied to the present data the results from earlier discriminant analyses which considered EP data for one semantic dimension (E, P or A) at a time. We found support for the possibility of using equations based on EP measures to establish reference coordinates in semantic space. These might be used to identify the semantic composition of more complex word stimuli.

(II) Representation of Numerical Meaning in Brain Potentials

The detection of numerical information by means of Evoked Potentials was explored by analyzing the responses to visual number symbols and a think cue (=). For both kinds of data the ten number classes (0-9) were discriminated at significantly better than chance levels by Discriminant Analyses using EP component scores as input variables. For both kinds of data, the discriminant functions developed for a group of subjects worked equally well for each individual.

(III) Storage in Short-Term Memory and Brain Responses

An Evoked Potential component with a post-stimulus peak about 250 msec. was found to be related to storage of information in short term memory. This Storage Component was found in an experiment investigating brain potentials in relation to an information processing task. In replications of this experiment at three different light intensity levels spaced 1.0 log unit apart, essentially the same component waveform and pattern of component scores were found. The memory storage interpretation was confirmed in a behavioral experiment which probed short-term memory. Recall was predicted by the magnitude of the Storage Component.

Multivariate Analysis of Evoked Potentials
and Semantic Meaning

The research is discussed below in the following parts:

(I) The Representation of Combinations of Connotative Meaning in Brain Potentials
  (A) Simultaneous Identification of Word Class and Scale Dimension
  (B) Separate Identification of Word Classes and Scale Dimensions
  (C) Reference Coordinates in Semantic Space Based on EP Analyses

(II) Representation of Numerical Meaning in Brain Potentials

(III) Storage in Short-Term Memory and Brain Responses

(I) The Representation of Combinations of Connotative Meaning in Brain Potentials

Our previously reported research indicated that EPs contain information about verbal, semantic meaning not dependent upon the particular word stimuli. Combinations of components of these EPs were powerful detectors of semantic differences. Such combinations also showed much promise in identifying the unknown semantic circumstances under which an EP occurred. This research has supported the feasibility of the general objective of inferring semantic meaning from analyses of brain waves. In the first phase internal semantic meaning was manipulated by carefully selecting stimulus words. In addition to internalized representations of semantic meaning elicited by stimulus words, another aspect of internalized representation may relate to an individual's semantic expectancies. When the same word is presented on different occasions, a subject may be seeking different kinds of semantic information. That is, a subject may have various kinds of semantic expectancies and, therefore, the semantic information in the words may be processed along various semantic dimensions. For example, an individual might be primarily concerned with potency (powerful-powerless) when a stimulus word "official" occurs, or he might be primarily concerned with evaluation (good-bad). Does the internal representation related to the word "official" vary for these different semantic expectancies? Do these different semantic expectancies have their own internal representations?

In order to study questions of this sort, we manipulated the semantic expectancy by assigning various semantic differential scales to the subjects at different times. The subject's task was the semantic differential task, as used by Osgood in developing his semantic analysis. This task requires giving each word a semantic differential rating on a designated scale. Different scales that are heavily loaded on (correlated with) each of the three Osgood dimensions: Evaluation (E), Potency (P), and Activity (A) were used (Table I).

In contrast to the work of earlier phases which involved the use of relatively "pure" semantic stimuli, the word stimuli used in this phase were semantically "complex." Instead of being defined in terms of a single connotative dimension, the meanings represented combintions of the evaluative, potency, and activity dimensions. The words were selected to be highly positive, negative or neutral on all three dimensions simultaneously.

Thus, basically a 3 X 3 factorial design was used: three semantic categories of words (representing high, neutral or low connotative values on all 3 dimensions) combined with 3 kinds of semantic differential tasks (pre-disposing the subject for semantic processing along the E, P, or A dimension). This permitted assessing the effect of the semantic meaning evoked by the words, the effect of the semantic set (context, expectancy) induced by the semantic differential task, and their interaction.

Synopsis of Procedure.

During each experimental run, 112 words were flashed in random order while the subject's EEG was recorded. For each run, there were 50, 50, and 12 words representing the High, Neutral and Low classes of semantic meaning lying along a diagonal of the Osgood dimensions: Evaluation, Potency, and Activity. The subject was assigned a particular semantic scale for use during the run in judging each word as it was presented. The EEGs for the stimulus words representing each semantic class were averaged for the run to obtain the evoked potentials (EPs) used in subsequent analyses. A total of 30 such runs were required to complete the collection of 90 such averaged EPs for each individual across all experimental conditions:

(1) Three semantic classes of stimulus words,
(2) Two replications
(3) Three semantic task dimensions, each represented by five different scales (to control for specific scale properties other than dominant semantic dimension).

TABLE I

Loadings of Semantic Differential Scales on Evaluation (E), Potency (P), and Activity (A) *

SCALE		E	P	A
E Dominantly				
E1	nice-awful	.96	-.02	-.09
E2	sweet-sour	.94	.02	-.04
E3	good-bad	.93	.03	-.05
E4	heavenly-unheavenly	.93	.00	-.21
E5	mild-harsh	.92	-.20	-.06
P Dominantly				
P1	big-little	-.05	.81	-.24
P2	powerful-powerless	.16	.75	.18
P3	deep-shallow	-.11	.69	-.32
P4	strong-weak	.04	.68	.13
P5	long-short	.02	.64	-.23
A Dominantly				
A1	fast-slow	-.14	.22	.64
A2	young-old	.39	-.42	.56
A3	noisy-quiet	-.39	.25	.56
A4	alive-dead	.52	.13	.55
A5	known-unknown	.16	.10	.48

----

* American English semantic differential loadings reported in Osgood, 1964. Loadings shown are for the first listed adjective of each pair. "Good", "Powerful", and "Fast" are represented by the positive poles of E, P, and A.

Details of Procedure

The research steps are summarized in the Flow Chart of Experiment (Table II).

Words with quantified semantic values on E, P, and A dimensions were selected from the available E, P, and A glossaries (Osgood, personal communication; Heise, 1971). We selected words which are semantically complex in the sense that they score high, neutral, or low on all three Osgood dimensions. Thus, three semantic meaning classes were used: High (E+,P+,A+), Neutral (E0,P0,A0), and Low (E-,P-,A-). The words were given in different random orders from run to run, so that the subjects could not anticipate either a semantic class or a particular word during the experimental runs.

Five scales that are heavily loaded on each of Osgood's three semantic dimensions (Evaluation, Potency, and Activity) were selected (Osgood, 1964). Each of these 15 semantic scales (Table I) was used with each stimulus word. This required 15 runs with Replicate 1 and 15 runs with Replicate 2, making a total of 30 runs for each subject. The scales were given in different random orders for each subject.

Before each run the subject was given the assigned semantic scale, e.g. "nice-awful," which he was to use on all 112 words in that run. The subject was asked to rate each stimulus word on the designated semantic scale using values from +3 to -3. The instructions to the subject when the scale was "nice-awful" were: "If the meaning of the word to you is more nice than awful, then give a + rating, with a 1, 2, or 3 to express various degrees of niceness. On the other hand, if the meaning of the word to you is more awful than nice, give a - rating using 1, 2, or 3 to indicate the degree of awfulness. If the word is perfectly neutral on that scale, give a "zero." If you felt that the word was very closely associated with one end of the scale, you might say "+3" or "-3." If you felt that the word was moderately associated with one or the other end of the scale, say "+2" or "-2." If the word seemed only slightly related to one side as opposed to the other, you might say "+1" or "-1." If you considered the scale completely irrelevant, or both sides equally associated, you should say "0." Make each item a separate and independent judgment." For each scale, regardless of whether it was "nice-awful," "big-little," "fast-slow," or some other scale, numerical values from +3 to -3 were used. After each word was flashed the subject gave his semantic differential rating verbally.

We have developed a computer-generated display system so that selected words can be individually presented to a subject as a briefly flashed stimulus on a CRT. The subject sat in a dark, sound-damped chamber. The average word subtended a visual angle of 1.5 degrees with a duration of 17 msec. Each letter was formed by lighting appropriate positions in a 5 by 7 matrix. A fixation

TABLE II

FLOW CHART OF EXPERIMENT

LISTS OF WORDS SELECTED
FOR 3 SEMANTIC CLASSES:
HIGH, NEUTRAL, LOW
ON ALL 3 OSGOOD
DIMENSIONS (E,P,A,)

----
5 SEMANTIC DIFFERENTIAL
SCALES SELECTED FOR EACH OF
3 DIMENSIONS: E, P, A
BASED ON OSGOOD'S ANALYSES

WORDS FLASHED ON CRT
EEG RECORDED
SUBJECT GIVES SEMANTIC DIFFERENTIAL

EVOKED POTENTIALS COLLECTED
FOR EACH SEMANTIC WORD CLASS
WITH EACH SEMANTIC SCALE

EPs STANDARDIZED WITHIN EACH OF 13 SUBJECTS
(MEANS = 0; S.D.s = 1)

VARIMAXED PRINCIPAL COMPONENTS ANALYSIS
ON EPs OF 102 TIME POINTS,
COMPONENT SCORES COMPUTED FOR EACH EP.

DISCRIMINANT ANALYSES USING
COMPONENT SCORES TO CLASSIFY
EPs INTO:

SEMANTIC
WORD
CLASSES (3)

SEMANTIC WORD CLASSES
AND SCALE DIMENSIONS
(3 x 3 = 9)

SEMANTIC
SCALE
DIMENSIONS (3)
----

target was presented (0.5 sec. duration) one second before each word. After each word was flashed the subject gave his semantic differential rating (+3 to -3) toward the end of the 2.5 sec. interval between each word and the fixation stimulus for the next trial. This task assured that each stimulus word was perceived and provided access to an important variable. The brain activity following these word stimuli was averaged separately for each of the semantic meaning classes in conjunction with each semantic scale. A computer program controlled the timing and delivered the stimuli and control pulses. The sequence for each word presentation (a trial) within each run was as follows:

(1) Fixation target on for 0.5 sec.
(2) Blackout for 0.5 sec.
(3) Stimulus word flashed (approximately 17 msec.)
(4) Blackout for 2.5 sec., during which time the subject gave a number representing his semantic judgment of the word on a designated scale.

A number of words (112) were presented in this fashion to constitute an experimental run. During experimental runs, the subject's EEG was being picked up from EEG electrodes, and recorded along with coded synchronization pulses associated with the various semantic word classes used.

Standard Grass electrodes (silver cup shape) were attached by bentonite CaCl paste. The analyses focused on a scalp location one-third of the distance from CZ to PZ (CPZ recorded monopolar to linked earlobes). The frequency bandpass of the recording system (Grass polygraph, FM tape recorder, operational amplifiers) was 0.1 to 70 Hz. Beginning with the word stimulus and lasting 510 msec., EPs were averaged by a program using 102 time points (5 msec. interval). Each EP was based on 50 or 12 different words of the same semantic class (50 for High and Neutral, 12 for Low). Eye movements were monitored with EOG (electrooculogram).

Data from 13 subjects are presented here. Each subject was given 30 runs of 112 words spread over a number of sessions.

The EP data from the various runs were collated in a manner and form suitable for multivariate statistical analyses. This involved disentangling the EP data from the random sequences, arranging them in a systematic order, and formatting them for paper tape and digital mag tape.

The data were standardized separately for each of the subjects. Using the BMDP1S Multipass Transgeneration Program (Dixon, 1975), each subject's data at each time point were transformed to z scores with means equal to 0 and standard deviations equal to 1. General advantages of preparing data for analysis in this way have been described by Rummel (1970, pp. 246-247). The specific reason for standardizing the data within subjects was to avoid swamping the semantic effects by individual differences in the subsequent analyses.

The principal components analysis closely followed the procedures which have worked well with our previous information processing data (Chapman, McCrary, Bragdon, and Chapman, in press). With those data the principal components analysis achieved a parsimonious representation of the data and the components were functionally meaningful. Two general steps are involved: (1) determing the EP components, and (2) measuring how much of each component is in each EP.

These steps were done by a varimaxed principal components analysis computed by BMDP4M Factor Analysis Program (Dixon, 1975). The EP data entered into the analysis were the intrasubject-standardized EP amplitude measurements obtained at the 102 successive time points for each of the EPs. The BMDP4M Program transformed the data matrix to a correlation matrix. The product-moment correlation coefficients computed for each pair of time periods comprised the 102 x 102 matrix to which principal component analysis was applied. Unities were retained in the diagonal. The number of components to be retained was set at the number of eigenvalues equal to or greater than unity. The retained components were rotated using the normalized varimax criterion (Kaiser, 1958). The analytic rotation preserves the orthogonality among the components while providing more distinct patterns, improving their clarity and definition. The varimaxed principal components method has performed well in achieving maximally parsimonious descriptions of a wide variety of data from differing scientific areas (Thorndike and Weiss, 1970) where other methods sometimes fail. Scores were computed for each of the original EPs on each of the varimaxed principal components. These component scores (factor scores, gain factors) measure the contributions of the components to the individual EPs. These component scores were compared for the various semantic classes of words.

Having reduced the dimensionality of the EP from 102 measures to a much smaller number of principal components, the next step was evaluating the extent to which these components contained semantic information and, more specifically, the utility of that information in discriminating and predicting semantic class of EPs. This evaluation was accomplished by multiple discriminant analyses. The aim of the discriminant analyses was to predict the semantic class membership of the EPs on the basis of the EP measures (component scores). The resulting discriminant functions are those which maximally separate the semantic classes. The discriminant analyses were done by the BMDP7M Stepwise Discriminant Analysis Program (Dixon, 1975). This program was applied to the component scores derived from the principal components analyses. A set of linear classification functions was computed by choosing the independent variables in a stepwise manner. Using these functions, the probabilities of each EP belonging to each semantic class was computed.

(A) SIMULTANEOUS IDENTIFICATION OF WORD CLASS AND SCALE DIMENSION

The basic experimental design employed makes it possible to examine two key questions involving semantically complex stimuli: (1) Can EP measures be used to determine simultaneously both the semantic class of words and the semantic dimension along which they were being judged by the subject? (2) Do the semantic meaning of the words and the semantic set induced by the task interact (do EP measures reflect different neural events for the stimulus word classes and the task scale dimensions)?

For the specific purpose of answering these questions, the EP data of each of the subjects were averaged to represent each of the 18 experimental combinations of 3 semantic classes of words X 3 scale dimensions X 2 replicates. These EPs were standardized at each time point (Mean = 0, S.D. = 1) separately for each subject. Varimaxed principal components analysis of the combined data of all 13 subjects (234 EPs X 102 time points) resulted in retaining 10 components accounting for 94.3% of the total variance. The scores for these components were used as the EP measures entered into discriminant analyses. In each analysis, discriminant functions were computed to distinguish among all 9 semantic conditions defined by the 3 semantic classes of words in combination with the 3 scale dimensions. The analyses were performed separately for the two different replicates in order to provide for cross-validations. Discriminant functions were obtained which detected statistically significant differences among the groups; the probabilities of the value of the F approximation to Wilk's lambda were less than .001 in each analysis. The usefulness of these functions was evaluated on the basis of the accuracy with which EPs could be assigned to the proper combinations of both word class and semantic scale. The results, combined for both of the word lists, are shown in Table III.

Since there are 9 groups to which an EP could be assigned, one out of 9 or 11.1% of the EPs would be expected to be correctly assigned by chance. The average apparent classification success rate obtained when clasifying the EPs used to develop the functions was 38%: 3 times better than chance.

The jackknifed cross-validation success rates estimate the outcomes expected if the classification functions were used to classify new EPs collected using the same list of words. While the overall average success rate shrinks to 22%, it remains 2 times better than chance.

The third part of Table III presents the results obtained when the classification functions were applied to data not used in their development and collected in the other replication. As might be expected the overall success rate is lowered. However, the 17% accuracy is higher than the percentage correct expected by chance and the overall chi-square test supports this difference as

TABLE III

9 Classes (3 Word Classes by 3 Scale Dimensions)

Percentage of EPs Correctly Classified

	Word Classes. High (E+,P+,A+)	Neutral (E0,P0,A0)	Low (E-,P-,A-)	Average
Development				
Scale Dimensions				
Evaluation	23.1	30.8	57.7	37.2
Potency	34.7	30.8	50.0	38.5
Activity	7.7	65.4	42.3	38.5
	----	----	----	----
Average	21.8	42.3	50.0	38.1
Jackknifed Cross-Validation				
Scale Dimensions				
Evaluation	7.7	15.4	50.0	24.4
Potency	30.8	7.7	23.1	20.5
Activity	0.0	46.2	19.3	21.8
	----	----	----	----
Average	12.8	23.1	30.8	22.2
Other-Replicate Cross-Validation				
Scale Dimensions				
Evaluation	19.3	7.7	15.4	14.1
Potency	15.4	19.3	23.1	19.3
Activity	7.7	34.6	11.6	18.0
	----	----	----	----
Average	14.1	20.5	16.7	17.1

Results combined for 2 Replicates; 13-subject group.

Each individual percentage based on 26 EPs.

Percentage correct expected by chance: 11.1%.

statistically reliable ($p < .005$).

The table reveals considerable variability in accuracy with which the 9 combinations are identified. Those combinations which involved the High (E+,P+,A+) class of words were detected less accurately than others, especially when semantic judgments about the words involve scales representing the Activity Dimension.

The use of the semantic differential task in conjunction with the three categories of words would be expected to predispose the subjects for semantic processing along the E, P or A dimension. We sought to assess the extent to which this would result in an interaction of semantic effects which would be represented in the EPs and influence the outcome of classifications. The 9 group discriminant analyses and classifications of the EPs (Table III) enable us to examine this question of the interrelationship of word classes and scale dimensions. This was statistically assessed by cross-tabulating the number of correct classifications in 3 X 3 contingency tables according to the semantic classes of the words and the semantic dimensions of the scales. Chi-square tests of independence give a mixed picture concerning independence of word class and subject task dimensions in determining classification outcomes. The independence hypothesis was rejected for the Development data ($p < .05$) and the Jackknifed Cross-Validation data ($p < .005$), but was accepted for the Other-Replicate Cross-Validation data (p between .10 and .25). It appears that there is a weak interaction between word class and scale dimensions, and that this interaction is weak enough that for first order approximation it is practical to treat their effects separately. We expect further analysis to clarify this question.

These analyses of classification data indicate that, as represented in the EP, the semantic processing of word stimuli and the set or processing imposed by a semantic task do not become greatly entangled. They do not interact sufficiently as to greatly influence (enhance or supress) the detectibility of one another.

(B) SEPARATE IDENTIFICATION OF WORD CLASSES AND SCALE DIMENSIONS

Since the analyses of the classifications above indicate that the effects in EPs related to distinguishing word classes are relatively independent of distinguishing semantic scale dimensions, separate classification functions were developed for each of these two kinds of semantic variables. The strategy was to compute discriminant analyses and develop classification functions for word classes and scale dimensions separately by entering the same data in both kinds of analyses but only specifying one or the other kind of group label while ignoring the other group label. The data entering these analyses were the same principal component scores that were used above in the simultaneous identification of word class and scale dimension (Table III). For the present purposes, however, the discriminant analyses were allowed to focus alone on either identification of word class or identification of scale dimension. To the extent that these two kinds of semantic variables have independent effects, the separately derived classification functions could be applied separately to the same EPs to "simultaneously" identify both word class and scale dimension, without loss of generality and perhaps with greater precision.

The results of separate identification of word classes and scale dimensions are summarized in Tables IV and V. For both kinds of analyses, separate discriminant analyses were made on the data obtained with the two replicates and the classification percentages averaged. For each of the discriminant analyses, discriminant functions were computed which detected statistically significant differences between the criterion groups. The chance probabilities of the F values computed from Wilk's lambda (U statistic) were less than .005 (most were less than .001). For both Tables IV and V the jackknifed and other-list cross-validations assess the success in applying the classification functions to data not used in their development: data obtained under the same conditions (one case left out) and data obtained by using the other replicate, respectively.

Separate identification of word class (Table IV) had an overall development success rate of 69%, which is to be compared with a chance rate of 33% (three word classes). The generality of the classification functions is indicated by the jackknifed cross-validation success rate (64%) and other-replicate cross-validation success rate (62%). These analyses indicate that word classes can be successfully identified in spite of the fact that a wide variety of semantic scales were being used by the subjects when these data were obtained.

Separate identification of scale dimension (Table V) had an average development success rate of 48%, which is to be compared with a chance rate of 33% (three scale dimensions). The generality of these classification functions is indicated by the jackknifed cross-validation (44%), but the other-replicate cross-validation is weak (33%). These analyses indicate that semantic scale dimensions can be successfully identified in spite of the fact that a wide variety of words were the specific stimuli for the EPs.

The identifications of word classes and semantic dimensions were not equally successful. The identification of the subject's task dimension (semantic scale) was not as robust and did not generalize to the other replication. The identification of the stimulus word class (High, Neutral, or Low) was quite robust and generalized strongly to the other replication (compare success rates of 69% and 62%).

In general, the separate identifications of word classes and scale dimensions (Tables IV and V) were significantly better than chance. It is to be noted that these success rates were obtained across subjects, i.e., the same classification functions were successfully used for all 13 subjects.

TABLE IV

3 SEMANTIC WORD CLASSES
MULTIDIMENSIONAL ANALYSIS AND CROSS-VALIDATION

Analysis ignores subject task: semantic differential ratings.

Percentages of EPs Correctly Classified

Semantic Class		Development Replicate 1	Jackknifed Cross-Validation Replicate 1	Other-Replicate Cross-Validation Replicate 2
High	(E+,P+,A+)	53.8	46.2	64.1
Neutral	(E0,P0,A0)	69.2	66.7	46.2
Low	(E-,P-,A-)	76.9	69.2	82.0
OVERALL		66.7	60.7	64.1
		Replicate 2	Replicate 2	Replicate 1
High	(E+,P+,A+)	76.9	71.8	53.8
Neutral	(E0,P0,A0)	61.5	56.4	56.4
Low	(E-,P-,A-)	76.9	71.8	66.7
OVERALL		71.8	66.7	59.0
COMBINED RESULTS		69.2	63.7	61.6

Each individual percentage based on 39 EPs

Percentage correct expected by chance: 33.3%

TABLE V

3 SEMANTIC SCALE DIMENSIONS
Analysis ignores word class.

Scale Dimension	Percentages of EPs Correctly Classified		
	Development Replicate 1	Jackknifed Cross-Validation Replicate 1	Other-Replicate Cross-Validation Replicate 2
Evaluation	46.2	38.5	33.3
Potency	59.0	56.4	28.2
Activity	48.7	46.2	33.3
	----	----	----
AVERAGE	51.3	47.0	31.6
	Replicate 2	Replicate 2	Replicate 1
Evaluation	43.6	43.6	28.2
Potency	41.0	35.9	12.8
Activity	48.7	43.6	64.1
	----	----	----
AVERAGE	44.4	41.0	35.0
	========	========	========
COMBINED RESULTS	47.8	44.0	33.3

Results obtained from 13-subject group.

Each individual percentage based on 39 EPs.

Percentage correct expected by chance: 33.3%.

(C) REFERENCE COORDINATES IN SEMANTIC SPACE BASED ON EP ANALYSES

The semantic glossary (Heise, 1971) tells us where the word stimuli we used are normatively located in semantic space as defined on the basis of Osgood's analyses. However, we wanted to assess, in at least a preliminary way, how the EPs in the present experiment might be located in semantic space using information developed from preceeding experiments on representation of semantic meaning in EPs.

A previous experiment, described in a recent report, investigated EP effects by presenting words representing six different classes of semantic meaning lying at the positive and negative extremes of each of the Osgood dimensions: Evaluation, Potency and Activity. We selected words which are relatively "pure" in the sense that, they score high or low on one of the dimensions and are relatively neutral on the other two. Twenty words from each of the six semantic categories (E+, E-, P+, P-, A+, and A-) were randomly assigned to a list. Two such lists were constructed using different words to control for specific stimulus characteristics or properties other than connotative meaning as well as provide a data base well suited for cross-validation. The other methodological features of the experiment, including the subjects' semantic differential tasks were essentially the same as those described for this current one. A part of the analysis of those data was directed at comparing results with still earlier studies and, for those purposes, a smaller EP data set was extracted ignoring the 15 semantic scales by averaging across them. For each of ten subjects, this resulted in EPs for six semantic classes for each of the two lists of words. The EP data were standardized separately for each subject (values at each time point brought to mean=0 and S.D.=1). The matrices of data for each subject were adjoined to form a 120 (EPs) by 102 (time points) input matrix for a varimaxed principal components analysis (Dixon, 1975). Eleven components exceeded the eigenvalue=1 criterion. Together these 11 components accounted for 93.9% of the variance. The scores for these components were used as the EP measures entered into discriminant analyses.

Six discriminant analyses were performed separately on the data from the three semantic dimensions (Evaluation, Potency and Activity) for the two "pure" word lists. In each of these "unidimensional" analyses, discriminant functions were computed which detected statistically significant differences between the two polar semantic groups. These differences were evaluated using the values of F computed from Wilk's lambda (U statistic). The chance probabilities of these F values were less than .05 to less than .001.

Overall, the unidimensional analyses of the "pure" semantic classes had an average apparent success of 94% and average jackknifed cross-validation success of 87%. It is to be noted that this success rate was obtained across subjects; the same classification functions were used for all ten subjects. When the same classification functions were applied to the EP data obtained from the other word list, the overall success rate was 74%. The A+ vs. A- classification functions largely contributed to the lowering of the success rates in cross-validation.

The EP data of the present experiment were also averaged across the 15 semantic scales and then standardized separately for each of the 13 subjects. Because the previous unidimensional analyses were on bipolar groups, only the EPs obtained using the High and Low stimulus words were included in the subsequent analyses. The matrices of data for each subject were adjoined forming a 52 (EPs) X 102 (time points) input matrix. Component scores for the 52 EPs were obtained using the component score coefficients obtained in the Principal Components Analysis of the 120 EPs in the previous experiments using "pure" words. The present data using "complex" words did not contribute to the Principal Components Analysis in any way. The classification functions previously obtained from the six unidimensional analyses were applied to the component scores for the EPs obtained with the "complex" semantic stimuli.

The success rates in classifying the semantically "complex" High and Low word classes using the predictor equations of the "pure" word classes were not far from chance levels overall. It was not expected that using E+ vs. E- prediction by itself, for example, would do well in placing words which are high or low on P and A dimensions as well. Greater classification successes would be expected from combining the information from all three types of predictor equations (E+ vs. E-; P+ vs. P-; A+ vs. A-). However, an important question to be answered in this development is whether there is any information derived from the previous "pure" discriminant functions which is applicable to the semantically "complex" words, and more specifically if these "simple" cannonical variates would place the "complex" word classes appropriatedly in semantic space. For example, the High word class being simultaneously E+,P+,A+ belongs on a diagonal between E+ and P+ and A+ axes defined by the "pure" cannonical variates.

Each of the canonical functions maximally separating the positive and negative "pure" word classes for each dimension was used, in turn, to compute coordinate values on E, P, and A canonical variate dimensions for the EPs obtained with the High and Low word classes. The mean coordinate values for the High and Low groups were statistically significantly (reliably) different from one another in the case of all six functions (p values ranged from $< .05$ to $<< .001$). In 5 out of the 6 analyses the High word class had mean values that were more positive than those for the Low word class. The function which produced means in the unexpected direction was one which faired poorly in discriminating A+ and A- groups in the earlier cross-validation.

In summary, we extracted the simplest set of reference coordinates available from our earlier experiments and used them to provide some estimates of the mean locations of EPs to more complex semantic stimuli. These preliminary results support further consideration of the possibility that EP measures provide an additional approach to mapping semantic space.

(II) Representation of Numerical Meaning in Brain Potentials

In addition to studying connotative meaning, we expanded the scope of the research to include investigations of denotative meaning and brain potentials. From the standpoint of experimental definition, design, and economy, number concepts have been selected for reasons similar to those which led us to the use of Osgood's analysis as a framework for examining connotative meaning. Number concepts provide us with objective, well-defined classes of meaning (i.e., sets containing a specific quantity) which have a variety of alternative physical representations (i.e., stimuli). For example, "3", "III", "THREE" and "three" all refer to the same number concept. These concepts also cut across most cultural boundaries and are not dependent on particular language groups. This research also represents an extension of our earlier work with numbers and evoked potentials (Chapman et al., 1964, 1965, 1966, 1969a, 1969b, 1973, 1974a, and in press). Extra-experimental reasons for selecting number concepts included the ubiquity and importance of number symbols in modern communications and transactions, and the related crucial importance of perceiving them correctly and understanding their meanings precisely. Number concepts are the linchpins of high speed man-machine-man interactions of an extremely wide variety. The analysis of the brain waves of the person receiving messages could inform the message sender, whether human or machine, if the messages had been correctly received.

Our present research combines into a single experimental design exploration of (1) the detection, by means of analysis of evoked potentials, of numerical information denoted by visual stimuli and (2) the transmission of such numerical information by means of brain potentials.

A set of cardinal numbers was used to define distinct categories of numerical meaning. Visual stimuli with various physical properties were used to represent each of these cardinal numbers, for example, "4", "IV", "four", "FOUR", etc. These stimuli were briefly presented visually on a computer controlled CRT display while the subject's EEG was recorded. The sequence of stimuli was: (1) a fixation symbol (*), (2) the number symbol, (3) the think cue symbol (=), and (4) the speak response cue (NOW). During the first two periods, the subject's task was to quietly observe the stimuli and to prepare for the response at the remaining periods. At the third period, the subject's task was to think the number as clearly as possible in synchrony with the cue stimulus, but not to vocalize its name at all. At the fourth period, the subject's task was to speak the number. EPs collected during the second period were studied for effects related to number concept and character set. Evoked potentials collected during the third period were used to determine whether numerical information can be transmitted by evoked potentials, without speech movements. Within each trial the onsets of the fixation, number symbol, think cue, and speak cue were 1.0 sec. apart. The numbers were presented in random sequences within each run.

Data were recorded from electrodes at scalp locations CPZ, OZ, C3, and C4. In addition, alpha EEG was automatically scored from OZ (Kropfl, Chapman & Armington, 1962) and EOG recorded from electrodes located infraorbitally and on the external canthus. All six channels of data were converted to digital values every 5 msec for a 1000 msec epoch beginning 25 msec before the stimuli. Dropping one time point at the time of the stimulus resulted in 199 time points in each Evoked Potential. On each trial, separate EPs were obtained to the number symbol and the think cue. EPs were averaged separately for each of the ten numbers (0-9) presented in each of 4 modes (Arabic, Roman, lower case, upper case) to both number and think cues. This resulted in 80 EPs (10 x 4 x 2). Each EP was the average of 54 trials (3 per run x 18 runs) obtained from a number of sessions. Data from five subjects were analyzed (400 EPs).

The EPs from each electrode were standardized separately for each subject (Chapman, McCrary, Chapman, & Bragdon, 1978). The standardization was accomplished by transforming the data at each time point to z-scores (mean=0, standard deviation=1). Next, the standardized data at each electrode were concatenated for all five subjects, forming separate data sets for each electrode (400 EPs). Each of these data sets was submitted to a separate Varimaxed Principal Components Analysis, using the correlation matrix of the 199 time points and eigenvalues=1 criterion. The resulting components accounted for more than 90% of the variance in each data set (Table VI). Component scores were computed for each of the EPs as part of the output of the Principal Components Analysis. The component scores were used as the EP measures in subsequent Discriminant Analyses, that investigated the relation of the brain responses to the experimental distinctions.

One of the experimental questions is whether the ten numbers could be discriminated by the EP responses to the number symbols, regardless of whether they were presented in Arabic, Roman, lower-case, or upper-case mode. A Discriminant Analysis was given access to the Component Scores from electrodes at CPZ, OZ, C3, C4 and alpha EEG in response to the number symbols and discriminant functions were computed to assign each of the 200 EPs to one of the ten number classes. Using ten of these component scores, the classification success on the data from which they were developed was 33.5%. The success rate expected by chance was 10%. The cross-validation assessed by the jackknifed procedure achieved a success rate of 19.5%. This is significantly better than chance (Chi-square=19.01, df=1, p<.0001). Thus, the EPs were significantly related to the number symbols, regardless of the four

Table VI

Principal Components Analyses for Number Experiment

Each EP contained 199 time points spaced 5 msec beginning 25 msec before stimulus. Each analysis based on 400 EPs (80 EPs x 5 subjects), each EP an average of 54 trials. EPs standardized within subjects separately at each electrode.

Electrode	Components	
	Number	Variance
CPZ	21	94.5%
OZ	19	91.7
C3	23	93.7
C4	21	94.0
Alpha	7	96.5
EOG	6	96.9

different modes with various physical properties. The success rates did not vary significantly for the Arabic, Roman, lower-case, and upper-case modes (Chi-square=3.15, df=3, p>.25). Nor did the success rates vary significantly among the five subjects (Chi-square=3.70, df=4, p>.25).

A second experimental question is whether the ten numbers could be discriminated by the EP responses to the think cue, which was always an equal sign (=). A Discriminant Analysis was computed on these equal-sign data in the same fashion as was done for the number symbols. Each of the 200 EPs to the equal-sign was assigned to one of the ten number classes according to the number that was to be thought. Using eleven of the EP component scores, the development classification success rate was 35.0%. The cross-validation assessed by the jackknifed procedure achieved a success rate of 21.5%. This was significantly better than the chance rate of 10% (corrected Chi-square=28.12, df=1, p<.0001). This success rate for EPs to the think cue was as good as the success rate for EPs to the number symbols themselves. The same discriminant functions performed equally well for all subjects (no difference in success rates among subjects, Chi-square=.37, df=4, p>.98).

In neither case did the discriminant functions generalize significantly between number symbol responses and think cue responses (9.0% and 13.0%). The EPs to the number symbols and the think cue (=) were quite distinguishable. Assigning the 400 EPs to these two classes was done with an accuracy of 96.5% (jackknifed success rate) by a Discriminant Analyses using 11 of the component scores. This was significantly better than the chance rate of 50% (corrected Chi-square=344.10, df=1, p<.0001).

(III) Storage in Short-Term Memory and Brain Responses

A critical ingredient in most, if not all, information processing by man is the temporary storage of incoming information so that it can be related and integrated in some fashion with other incoming information. There is considerable behavioral evidence that all incoming stimulus information is not equally available at a later time (beginning with Ebbinghaus's classic studies of learning and memory in the ninteenth century). There are a number of factors which influence the storage and retention of stimulus information in memory. Among these are the relevance of the stimulus information to the person's task at the moment and the amount of previous information being retained in a person's short-term memory which is apparently of limited capacity. Regardless of the reasons why stimulus information sometimes is not stored in an individual's memory, it would be extremely useful from a theoretical and practical standpoint to be able to determine when stimulus information is or is not being stored in memory.

We have described an EP component which is tentatively interpreted as being related to information storage (Chapman, 1974a; Chapman, McCrary, Bragdon, and Chapman, in press). It was found in an experiment which was investigating brain potentials in relation to a number/letter information processing task. A number of EP components were identified which were functionally related to various features of information processing. Among these is one which appears to be quite specifically related to information storage and which appears in the same form in replications of the experiments in which the intensity of the stimulus was varied across 2.0 log units. The experiment used number and letter comparison tasks in which subjects performed different information-processing operations on different occurrences of the same physical stimuli (Chapman, 1973). On each trial, four stimuli, two numbers and two letters, were flashed in random order with an interval of 3/4 sec. For some trials, the subject indicated whether the first or second number was larger by appropriately moving a two-way switch, the letters being irrelevant. For other trials, the subject compared the letters and indicated the alphabetic order. Since the numbers and letters were randomly selected (1-6, A-F), the sequences of numbers and letters were randomized, and the performance accuracy was better than 99%, nearly every stimulus was processed appropriately by the subjects. This task required the subject to store the letter and/or number information in his memory in order to compare it with relevant stimulus information occurring a short while later. An EP component was found by a Principal Components Analysis which was associated with storage of stimulus information. The component scores for this "storage" component were relatively high for relevant and irrelevant stimuli in the first of the intra-trial positions and for relevant stimuli in the second position. This was in marked contrast to the "storage" component scores to the remaining stimuli where it was not necessary for the subject to store the stimulus information either because it was irrelevant to the task or could be used immediately to compare with the memory of the previous relevant stimuli. It was interpreted that the stimulus information was stored whenever it was the first relevant stimulus or when the memory capacity was not being taxed by holding previously stored information. The loadings for this "storage" component reached their maximum at about 250 msec. This "storage" component is orthogonal to other EP components found in the same experiment, such as P300 and CNV-resolution. Furthermore, the pattern of component scores for this "storage" component is different and more specifically related to information storage than P300, which was related to all relevant stimuli, or CNV-resolution, which was related to stimulus uncertainty. Essentially the same data in terms of both pattern of component scores and time-course of component loadings have been found in replications of this experiment at 3 different intensity levels spaced 1.0 log units apart. This lends confidence to the potential generality of the "storage" component.

The waveforms of these components are very similar in all four sets of data, reaching their maximum about 250 msec. after the stimulus. The coefficients of factorial similarity among the waveforms from the four data sets were high, ranging between 0.85 and 0.99. For the previous experiment the maximum was at 250 msec.; for the new data the maxima are at 250 msec., 250 msec., and 270 msec. for high, mid, and low light intensities, respectively.

The Storage Component tends to be positive for stimuli whose information needs to be stored by the subject (Fig. 1). Thus, the magnitude of the Storage Component was more positive for the first of the two relevant stimuli presented on each trial (intra-trial positions 1 or 2) than for the second relevant stimulus (intra-trial positions 3 or 4). The Storage Component was also relatively positive for the irrelevant stimuli when they occurred in intra-trial position 1. Extending the storage interpretation to this result leads to the hypothesis that an irrelevant stimulus in position 1 is stored in memory, whereas irrelevant stimuli in positions 2, 3, and 4 are not. This may be related to short-term memory having a limited capacity and storage of irrelevant information interfering with processing relevant information. The difference in the Storage Component scores for relevant and irrelevant stimuli in intra-trial position 2 is evidence that this component is not related simply to an order effect. Nor is this EP component related to amount of processing which is presumably greatest for the comparison operations following the second relevant stimulus, next most for the storage operations associated with the first relevant stimulus, and least for the irrelevant stimuli. Nor does this EP component reflect a general relevant-irrelevant distinction. The Storage Component did not consistently distinguish between number and letter processing, or between number and letter stimuli. The simplest and most direct interpretation is that this EP component is related to the storage of information in the subject's short-term memory. More specifically the component may reflect the process of reading information out of a sensory register into short-term memory. Not

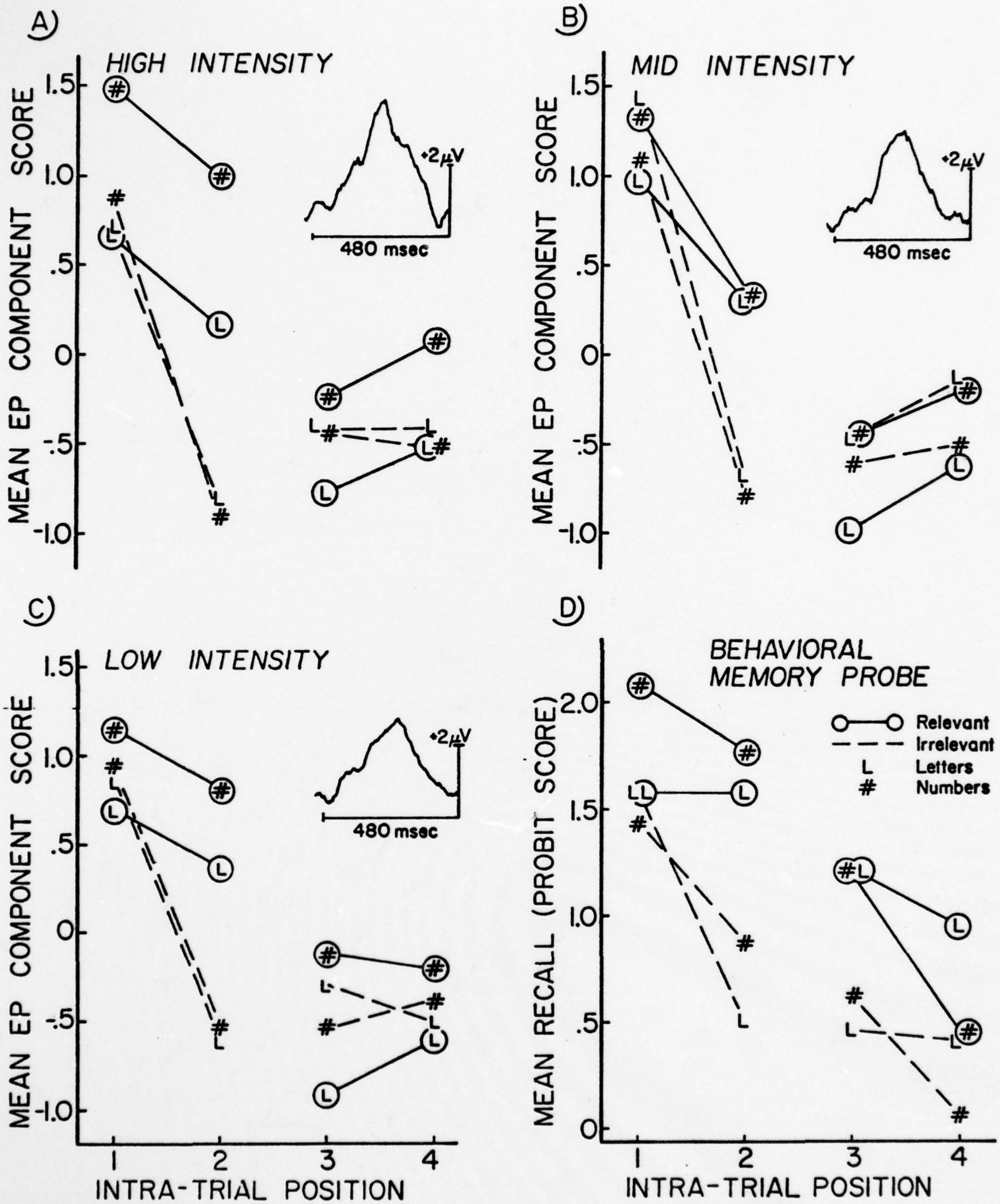


Fig. 1 Storage Component of EPs from CPZ and recall for experimental conditions in which short-term memory demands vary. Insets show the Storage Component waveforms scaled appropriately for relevant numbers in position 1.

only were the Storage Component scores related to memory storage conditions, but also the timing of this EP component (waveforms in Fig. 1 insets) is appropriate for information storage. The maximum of the Storage Component was at 250 msec. This is an appropriate time for storing information needed later since the literature suggests that the sensory register (icon) is fading about that time.

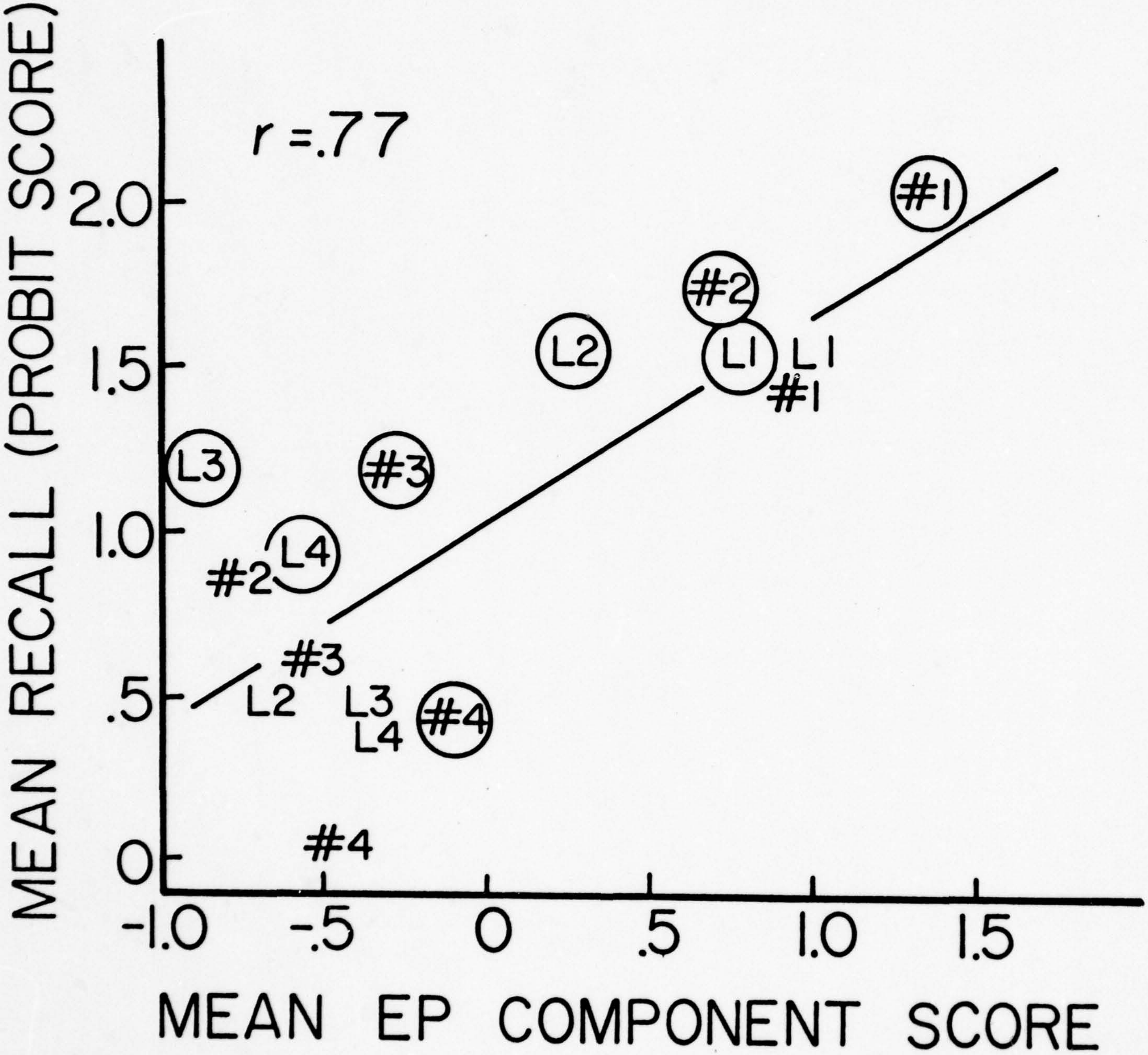
The results demonstrate the robustness of the Storage Component in the face of large differences in the physical parameters of the stimuli. That the Storage Component represents neural activity in the stimulus-response sequence that occurs later than the simple processing of sensory input is supported by two findings: (i) its independence of whether the stimuli are numbers or letters and (ii) that changes in stimulus intensity which are sufficient to alter markedly the overall EP have only a small effect on the Storage Component. Further, that the Storage Component occurs after the simple processing of sensory input, including recognition of the informational content of the stimulus, is indicated by the differences in response to identical physical stimuli when they play different roles in the information processing task.

Our tentative interpretation that this EP component is related to storage was based on considering the differential scores for the 1st and 2nd relevant stimuli within each trial. However, finding high Storage Component scores for irrelevant stimuli in position 1 required ad hoc interpretations in order to maintain the storage identification. Therefore, we felt that it was important to check more directly the storage interpretation by a behavioral experiment designed to assess storage in short-term memory.

The behavioral experiment used a memory probe technique to test the subjects' recall of individual stimuli for each of the 16 conditions in the electrophysiological experiments. Experimental sessions and data collection were conducted by experimenters not involved in the previous experiments and not aware of the hypothesis being tested. The experimental procedure was the same as for the collection of brain responses with the addition of occasional memory probes. The primary task on each trial was to compare the two numbers on one run of 102 trials and to compare the two letters on a second run. Within each run of 102 trials, eight randomly located memory probes were selected to test recall of a letter and a number in each of the four intra-trial positions. Without prior warning of when probes would occur, blank flashes were delivered 3/4 and 1 1/2 sec. after the probed stimulus and the subject was asked what the last character was. These blank flashes were used to mask the probed stimulus and to delay the recall report in order to reduce the effects of very short-term sensory register. From each subject, one such recall probe was obtained for each of the 16 conditions (8 probes each in a number-relevant and a letter-relevant run). The percent correct recalls from 52 subjects (29 female and 23 male college students) are given in z-score units in Fig. 1D. The pattern of correct recalls is strikingly similar to the pattern of Storage Component scores in Fig. 1A, 1B, 1C with better memory for relevant stimuli in intra-trial positions 1 and 2 and irrelevant stimuli in position 1. The six correlations among the four patterns of 16 means each in Fig. 1 ranged from .71 to .97. Three interesting features of the data are common to both the Storage Component of the brain responses and the subjects' short-term memory: (1) the first relevant stimulus on a trial (intra-trial position 1 or 2) gave higher scores than did the second relevant stimulus (positions 3 or 4); (2) the scores were high for both relevant and irrelevant stimuli in intra-trial position 1; and (3) in position 2, the scores were higher for relevant stimuli than irrelevant stimuli. The recall performance is plotted as a function of mean Storage Component score (averaged over the three intensity levels) in Fig. 2. Thus, the storage interpretation was confirmed by predicting recall performance on the basis of the Storage Component of brain responses ($r = .77$). The accuracy of this prediction is impressive considering that behavioral recall is not solely a function of storage but is generally considered to be greatly influenced by other factors including retrieval mechanisms.

One of the reasons that the Storage Component has not been found in other EP research is that it may be partially masked by a positive peak in the EP which often occurs slightly before 250 msec. Hence, measurement based on peaks of the average EP may miss the latent Storage Component that was derived by Principal Component Analyses which assess the relationships among all the time points and decompose EPs into independent sources of variation. Now that the Storage Component has been described and its waveform is known, it may be measured by computing component scores directly in other EP studies without doing a complete Principal Component Analysis.

Fig. 2 Behavioral recall as a function of brain response Storage Component score. Experimental conditions: letters (L) or numbers (#); intra-trial positions (1, 2, 3, or 4); relevant to primary task (circled) or irrelevant (not circled).



CONCLUSIONS

Using linear combinations of EP component measures, statistically significant differences were found among the pre-defined semantic groups in all analyses. The magnitude of the differences among the semantic groups in the samples is illustrated by the success of these linear combinations in classifying the EPs from which they were computed. In the analyses of word classes (Table IV), the success rate was about 2 times greater than what one would expect (on the average) if assignments were random or no true differences existed among the groups. It is concluded that evoked potential component measures can distinguish differences among "complex" word classes which have been defined by means of Osgood's dimensions in semantic space.

The results of these analyses of the data provide additional confirmation of our previous findings: EPs contain information about the semantic meaning of word stimuli used to obtain them and that combinations of components show promise in identifying the unknown semantic character of stimuli which have evoked particular brain responses.

The results obtained in the current phase further indicate that semantic effects in the EP continue to be detectible when the subject is engaged in a semantic task considerably more complex than only repeating the stimulus words. The added complexity of the experimental conditions clearly does not obscure the stimulus effects.

Our new findings extend the generality of detecting semantic word classes to three more connotative meaning classes: High, Neutral, and Low. These were defined in terms of Osgood's E,P,& A dimensions, being simultaneously high, neutral, or low on all three semantic dimensions. Some preliminary computations also indicate the possibility of using EP measures to develop reference coordinates which can be used to specify relative locations in semantic space for more complex semantic stimuli.

In a manner which parallels our conclusions about identifying stimulus word class, there is some generality to identifying scale dimension. A number of scales were used to represent each semantic dimension (five for each) in order to establish general relationships to EPs, not tied to particular exemplars of the semantic scales. This parallels the use of many exemplars of stimulus word class in establishing the generality of those EP effects. However, identifying the scale dimension used by the subjects is not as robust as identifying the semantic class of the stimulus word.

These findings have implications for applications as well as a basic understanding of the processes. Two kinds of semantic effects are registered in the EP and can be used to tap different aspects: (1) assessing the processing of the semantic meaning in the words, regardless of the semantic expectancies of the subject, and (2) assessing the semantic expectancies of the subject, regardless of the semantic content of the words.

The importance of these semantic effects lies partly in that they may assess communication at a very high level of understanding. Connotative meaning is closer to the reactive side of information processing than to the input side. The connotative dimensions reflect general characteristics of the referents of the incoming information: is it good or bad (Evaluation Dimension); is it powerful or weak (Potency Dimension); is it active or passive (Activity Dimension). Such overall assessment of incoming information has two important implications. One implication is that appropriate connotative responses depend on appropriate processing of the incoming message and thus may be used to assess understanding at a rather high level. The other implication is that the connotative responses ought to be a fairly good predictor of actual behavior in relation to the information, regardless of whether such behavior is appropriate. For example, if information about an "angry lion nearby" were responded to with connotative responses of "good", "weak", and "passive", then regardless of the source of misunderstanding, it would be a good bet that the receiver with such connotative responses may not behave in appropriate ways.

The detection of numerical information by means of Evoked Potentials was explored by analyzing the responses to visual number symbols and a think cue (=). For both kinds of data the ten number classes (0-9) were discriminated at significantly better than chance levels by Discriminant Analyses using EP component scores as input variables. Jackknifed cross-validation classification success rates were approximately twice as large as expected by chance. This probably represents the first data relating brain response data to numerical concepts. For both kinds of data, the discriminant functions developed for a group of subjects worked equally well for each individual. The research has established the feasibility of identifying differences among number concepts by analysis of brain potentials, and, thus, pending careful confirmation, opens the way not only to further studies of numerical concepts but also to research involving the manipulation of numbers.

An Evoked Potential component with a post-stimulus peak about 250 msec. is related to storage of information in short-term memory. This Storage Component was found in an experiment investigating brain potentials in relation to a number/letter information processing task. In replications of this experiment at three different light intensity levels spaced 1.0 log unit apart, essentially the same component waveform and pattern of component scores were found. The memory storage interpretation was confirmed in a behavioral experiment which probed short-term memory. Recall was predicted by the magnitude of the Storage Component.

Regardless of the reasons why stimulus information sometimes is not stored in an individual's memory, it would be extremely useful from theoretical, experimental, educational and clinical standpoints to be able to determine whether or not stimulus information is being stored in memory. If further research sustains the interpretation that the Storage Component of Evoked Potentials reflects the process of storing information in short-term memory, then this brain response component may be used to assess storage per se, uncontaminated by retrieval mechanisms. The Storage Component of Evoked Potentials holds promise of serving this practical function as well as providing an entry to understanding the neural processes related to memory.

REFERENCES

Chapman, R.M. Discussion on eye movements, CNV and AEP. In Donchin, E. and Lindsley, D.B. (Eds.) Averaged evoked potentials: Methods, results, and evaluations. Washington, D.C., NASA SP-191, 1969a, 177-180.

Chapman, R.M. Definition and measurement of "psychological" independent variables in an average evoked potential experiment. In Donchin, E. and Lindsley, D.B. (Eds.) Average evoked potentials: Methods, results, and evaluations. Washington, D.C., NASA SP-191, 1969b, 262-275.

Chapman, R.M. Evoked potentials of the brain related to thinking. In McGuigan, F.J. and Schoonover, R. (Eds.) Psychophysiology of Thinking. New York: Academic Press, 1973, 69-108.

Chapman, R.M. Latent components of average evoked brain response functionally related to information processing. In International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts. Brussels: Presses Universitaires de Bruxelles, 1974a, 38-42.

Chapman, R.M. Semantic meaning of words and average evoked potentials. In International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts. Brussels: Presses Universitaires de Bruxelles, 1974b, 43-45.

Chapman, R.M. Light intensity and the "storage" component in visual evoked potentials. Electroencephalography and Clinical Neurophysiology, 1977, 43, 778.

Chapman, Robert M. Language and evoked potentials. In D.A. Otto (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. U.S. Government Printing Office, Washington D.C., EPA-600/9-77-043, in press.

Chapman, Robert M. Method of EP Analysis in Linguistic Research. In D.A. Otto (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. U.S. Government Printing Office, Washington D.C., EPA-600/9-77-043, in press.

Chapman, Robert M. Connotative Meaning and Averaged Evoked Potentials. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior. New York: Plenum Press, in press.

Chapman, R.M. On the road to specific information in evoked potentials. In Armington, J.C., Krauskopf, J. and Wooten, B. (Eds.), Visual Psychophysics: Its Physiological Basis. New York: Academic Press, in press, 469-480.

Chapman, R.M. and Bragdon, H.R. Evoked responses to numerical and non-numerical visual stimuli while problem solving. Nature, 1964, 203, 1155-1157.

Chapman, R.M., Bragdon, H.R., Chapman, J.A. and McCrary, J.W. Semantic meaning of words and average evoked potentials. In J.E. Desmedt (Ed.), Progress in Clinical Neurophysiology. Vol. 3: Language and Hemispheric Specialization in Man: Cerebral Event-Related Potentials. Basel: Karger, 1977, 36-47.

Chapman, R.M. and Chapman, J.A. The General Automation 18/30 as a system for the general analysis and acquisition of data in physiological psychology. Behav. Res. Methods and Instrum., 1972, 4, 77-81.

Chapman, R.M. and McCrary, J.W. Hemispheric differences in evoked potentials to relevant and irrelevant visual stimuli. In E. Callaway and D. Lehmann (Eds.), Event Related Potentials in Man: Applications and Problems. New York: Plenum Press, in press.

Chapman, R.M., McCrary, J.W., Bragdon, H.R. and Chapman, J.A. Latent components of event-related potentials functionally related to information processing. In J.E. Desmedt (Ed.) Progress in Clinical Neurophysiology. Vol. 6: Cognitive Components in Cerebral Event-related Potentials and Selective Attention. Basel: Karger, in press.

Chapman, R.M, McCrary, J.W., Chapman, J.A. and Bragdon, H.R. Brain responses related to semantic meaning. Brain and Language, 1978, 5, 195-205.

Chapman, R.M., McCrary, J.W. and Chapman, J.A. Short-term memory: The "storage" component of human brain responses predicts recall. Science, 1978 in press.

Dixon, W.J. (Ed.) BMDP Biomedical Computer Programs. Berkeley: University of California Press, 1975.

Heise, D. R. Evaluation, potency, and activity scores for 1551 words: A merging of three published lists. Dept. of Sociology, University of North Carolina, Chapel Hill (1971).

Kropfl, W.J., Chapman, R.M. and Armington, J.C. Apparatus for scoring selected electroencephalographic rhythms. Electroencephalography and Clinical Neurophysiology, 1962, 14, 921-923.

Osgood, C.E. Semantic differential technique in the comparative study of cultures. American Anthropologist, 1964, 66, 171-200.

Rummel, R.J. Applied factor analysis. Evanston: Northwestern Univ. Press, 1970.

Sheatz, G.C. and Chapman, R.M. Task relevance and auditory evoked responses. Electroencephalography and Clinical Neurophysiology, 1969, 26, 468-475.

BRAIN AND LANGUAGE 5, 195–205 (1978)

Brain Responses Related to Semantic Meaning

ROBERT M. CHAPMAN, JOHN W. MCCRARY, JOHN A. CHAPMAN, AND HENRY R. BRAGDON

University of Rochester

Evoked Potentials from electroencephalogram (EEG) recording were averaged to many visually presented word stimuli whose semantic meanings were specified along Osgood's semantic dimensions of Evaluation, Potency, and Activity [Miron & Osgood, 1966, in R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*, Chicago: Rand–McNally; Osgood, 1971, *Journal of Social Issues*, **27**, 5–63; Osgood, May, & Miron, 1975, *Cross-cultural universals of affective meaning*, Urbana, Ill.: University of Illinois Press]. Multivariate analyses classified the Evoked Potentials to six semantic classes with success rates more than twice chance expectation. The pattern of brain activity related to the six semantic classes was similar for (i) two sets of words, (ii) 10 subjects used to develop the analyses, and (iii) an added, new subject.

The technique of averaging electrical activity from the human scalp to repeated, discrete sensory, motor, and cognitive events has found extensive use in extracting brain responses (Evoked Potentials or EPs) from the ongoing electroencephalogram (EEG) (for reviews: Callaway, 1975; Regan, 1972). In order to investigate brain responses related to semantic meaning, we extended this technique to averaging EPs across a number of words belonging to the same semantic class (Chapman, 1974b; Chapman, Bragdon, Chapman, & McCrary, 1977). With the aid of a quantified theory of connotative semantic meaning and multivariate statistical techniques, we found brain activity from the human scalp which is related to semantic meaning. Our study focuses on *intralinguistic* differences, that is, on brain response waveform differences related to distinctions within a particular language domain, namely connotative meaning. This strategy strengthens the interpretation of the positive results in that (i) classes of variables such as stimulus differences, general state differences, and information processing differences are less likely to confound the result and (ii) the specificity of the language effects is more

Supported in part by NIH Research Grant No. 5 R01 EY01593 from the National Eye Institute, the Advanced Research Projects Agency (monitored by ONR under Contract N00014-76-C-0185), and a grant from the Eye Research Foundation of Bethesda, Maryland. Send reprint requests to Robert M. Chapman, Center for Visual Science, Psychology Department, University of Rochester, Rochester, New York 14627.

0093-934X/78/0052-0195$02.00/0

striking. In order to control commonly confounding variables, the subject's task was held constant, the presentation sequences were randomized, and the semantic classes were represented by a relatively large number of different words in two lists. With regard to the specificity of the linguistic effects, six different semantic classes were distinguished.

MATERIALS AND METHODS

We specified and controlled internal semantic meanings using the conceptions and materials provided by Osgood's *analyses of semantic meaning* (Miron & Osgood, 1966; Osgood, 1971; Osgood, May, & Miron, 1975). Those analyses indicate that the connotative meaning of a word may be represented by its position in a space spanned by three semantic dimensions: Evaluation, Potency, and Activity (E, P, and A). Such quantitative descriptions of words in a three-dimensional semantic space allow similar words to be selected by those characteristics to form semantic classes. We selected words (Heise, 1971) which are relatively "pure" in the sense that they score high or low on one of the dimensions and are relatively neutral on the other two. Thus, we used six semantic meaning classes (E+, E−, P+, P−, A+, A−) representing the positive and negative extremes of the Evaluation, Potency, and Activity dimensions. Average values for the words were 2.0 and −1.3 for the E+ and E− classes, 1.9 and −.6 for the P+ and P− classes, and 1.0 and −.8 for the A+ and A− classes on their respective semantic dimensions (range = +3 to −3). Some of the words in each of the semantic classes are: E+, peace, food, pleasure, fresh; E−, greed, thief, bitter, lizard; P+, judge, hard, ocean, official; P−, jelly, little, feather, five; A+, tennis, surprise, worker, spicy; and A−, silence, stone, poetry, past.

Twenty words from each of the six semantic classes constituted a list. Two such lists were randomly constructed using a total of 220 different words. The same words were used in both lists for the P− class due to their scarcity in Heise's compilation. The 120 different words within each list were given in different random orders from run to run, so that the subjects could not anticipate either a semantic class or a particular word during the experimental runs. Each word was visually presented on a computer-controlled CRT display while the subject's EEG was recorded.

The sequence for each word presentation (a trial) within each run was: (i) fixation asterisk on for .5 sec, (ii) blackout for .5 sec, (iii) stimulus word flashed (about 17 msec by average photocell measurement), and (iv) blackout for 2.5 sec, toward the end of which time the subject said the word. This simple task of saying each word assured that it was perceived. The average word was 1.5° visual angle, 17 msec in duration, and about 1.4 log units luminance above the word recognition threshold.

Statistical analyses were made of the stimulus words in relation to their luminance and luminance distribution. The stimuli were composed of dot-matrix characters, so the relative luminance flux could be measured by the relative number of dots in each word. When this measure was statistically analyzed for its ability to discriminate the six word classes, it resulted in an overall classification success of only 19.6 per cent, which was not significantly different ($.20 < p < .30$) from the chance level of 16.7%. The possibility that the luminance distribution might be different for words in the various semantic classes was assessed by comparing letter frequencies. The distribution of letters was not significantly different for the six word classes. This was assessed by counting the number of times each of the 26 letters occurred in each word of each semantic class represented in two different lists of words and entering these values into an analysis of variance. As expected, the letters differed significantly in their overall frequencies of occurrence in the words as a whole (e.g., "E" was quite frequent and "X" was infrequent). These differences, however, did not change systematically according to the semantic class of the words. There was no significant interaction with semantic class or with word list.

The subjects were run individually in a dark, sound-damped chamber to reduce contamination from eye movements and sounds. Subjects sat and directly viewed the CRT display at a 1-m distance within the otherwise dark chamber.

The data reported here were recorded monopolar between a scalp electrode located on the midline over the central-parietal area (CPZ) and linked earlobes. The scalp electrode was a Grass electrode (silver, cup shaped) attached by bentonite CaCl paste one-third of the distance from CZ to PZ (International 10–20 System). Grass ear electrode clips on the earlobes served as reference electrodes and were electrically connected at the input to the amplifier. The electrical potentials were amplified by Grass EEG amplifiers and monitored on a Grass polygraph. The data were recorded on an AMPEX FM tape recorder and later fed to a computer via an interface containing operational amplifiers. The frequency bandpass of the overall recording system was .1 to 70 Hz (half amplitude). Eye movements were monitored by bipolar EOG recording using another recording channel (the same gain and frequency response).

The 10 unselected subjects were female (six) and male (four) paid volunteers. Their ages were 18–23 years and their educational background varied considerably (typically high school graduates). None appeared to know about Osgood's analyses.

Beginning with the stimulus word and lasting 510 msec, EPs for each semantic class were averaged from the EEG by a program using 102 time points (5-msec interval). Each EP was averaged over 20 different words of the same semantic class. Thus, EPs were collected for each of the six semantic classes for each of the two word lists. The fundamental idea is that the neural components which are common to the words of a given semantic meaning class will appear in the average EP, while those aspects of brain processing which are not common will tend to cancel out.

Over a number of sessions, each subject was given 12 to 20 runs of 120 words (20 words in each of six semantic meaning classes) and equal numbers of List 1 and List 2 runs were randomly interspersed. Due to scheduling problems different subjects received different numbers of runs. However, the design was balanced for each subject (word classes within runs and lists within sessions). Twelve EPs (six semantic classes for List 1 and for List 2) were obtained for each subject by further averaging the appropriate EPs from the available runs. Thus, the analyses were based on EPs averaged from 6 to 10 runs for various subjects ($N = 120$ to 200 for each semantic class for each list).

RESULTS

Summary data from 10 subjects are presented here. Although the overall average brain responses to the six semantic classes (the left part of Fig. 1) appear very similar to each other, the small differences associated with the semantic classes were consistent for both lists and all 10 subjects. The same data after standardizing separately for each subject (see below) and then averaging (the right part of Fig. 1) clearly show different waveforms for the different semantic classes.

Our data analyses involved: (i) standardizing the EPs for each subject separately, (ii) computing a Principal Components Analysis using all of the standardized data for the 10 subjects, and (iii) computing Multiple Discriminant Analyses using the component scores obtained from the Principal Components Analysis as the input variables and the six semantic classes as the criterion variables.

The data were standardized separately for each of the subjects. Using the BMDP1S Multipass Transgeneration Program (Dixon, 1975: program

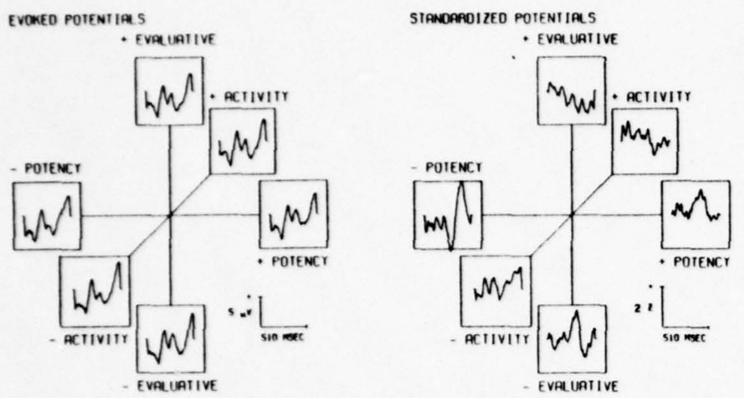


FIG. 1. Average brain responses for six semantic classes before and after standardization. The semantic classes are based on Osgood's Evaluation, Potency, and Activity dimensions (Miron & Osgood, 1966; Osgood, 1971; Osgood et al., 1975) which define a three-dimensional semantic space, represented schematically here. The EPs cover 510 msec (102 time points × 5 msec) along the horizontal axis, beginning at the time the words were flashed. The vertical axes for the EPs are in microvolts, in the left panel, and are in standard units, z, in the right panel. For the Standardized Potentials, each subject's data at each time point were transformed to z scores (*means* = 0 *and standard deviations* = 1). Averages include data for two lists and 10 subjects. Monopolar recordings (bandpass: .1 to 70 Hz) are from a scalp location one-third of the distance from Cz to Pz.

revised Oct. 7, 1974), each subject's data at each time point were transformed to z scores with means equal to 0 and standard deviations equal to 1. The general advantages of preparing data for analysis in this way have been described by Rummel (1970, pp. 246–247). The specific reason for standardizing the data within the subjects was to avoid swamping the semantic effects by individual differences in the subsequent analyses.

A Principal Components Analysis (Dixon, 1975; Kaiser, 1958), which followed procedures previously described (Chapman, 1974a; Chapman, McCrary, Bragdon, & Chapman, in press), was computed in order to: (i) determine the EP components and (ii) measure how much of each component is in each EP. The data entering the analysis were 120 EPs (6 semantic classes × 2 lists × 10 subjects) measured at 102 time points. The options used included: correlation matrix with unities in diagonal, eigenvalue = 1 as the cutoff, and rotation to the normalized varimax criterion (Kaiser, 1958). The 12 retained components (eigenvalues > 1) accounted for 94% of the variance. Scores measuring the contributions of the 12 components to the individual EPs were computed.

Having reduced each EP from 102 measures to only 12, the next step was evaluating the extent to which these 12 brain response components contained semantic information. This was done by Multiple Discriminant Analyses (Dixon, 1975), the aim of which was to predict the semantic class

membership of the EPs on the basis of the EP measures. A set of linear classification functions was computed by the program choosing the EP components according to how well they discriminated among the semantic classes. Using these classification functions, each EP was assigned to one of the six semantic classes.

Discriminant classification analyses using the evoked potential component scores to distinguish semantic classes were of two kinds: (1) unidimensional and (2) multidimensional. The unidimensional analyses considered the data for one semantic dimension at a time (E, P, or A), in which case two semantic classes (positive and negative classes from one dimension) were discriminated. The multidimensional analyses considered the data for all three semantic dimensions at once, in which case six semantic classes were discriminated from each other (positive and negative extremes of E, P, and A dimensions).

Classification functions were developed separately for the data from each list of words and the results were crossvalidated by several procedures: (i) jackknifed, (ii) other-list, and (iii) fresh data of a new subject. The jackknifed crossvalidation is used to estimate the success which would be expected in classifying other, additional EPs obtained using the development list. In the other-list crossvalidation, the classification rule developed for EPs obtained with one word list is used to classify EPs collected with the other list of word stimuli. This provides a further check on generalizability of the classification functions and tests their likely success rate in classifying other, additional EPs obtained using a different set of words.

The results concentrate upon evaluating the usefulness of EP components in distinguishing among semantic classes and, thus, are reported primarily in terms of EP classification success rates. The extensive tables of the intermediate computational results (rotated component loadings, component score coefficients, and coefficients for 24 group classification functions) are available upon request to the principal author.

Unidimensional Analyses of Semantic Groups

Six discriminant analyses were performed separately on the data from the three semantic dimensions (Evaluation, Potency, and Activity) for the two word lists (List 1 and List 2). In each of the analyses, classification functions were computed which detected statistically significant differences between the groups. These differences were evaluated using the values of F computed from Wilk's λ (U statistic). The chance probabilities of these F values ranged from less than .01 (Evaluation dimension, List 2) to less than .001 (Potency dimension, both lists).

The results obtained when these functions were used to classify EPs into semantic groups are summarized in Table 1. For example, the results for the Evaluation semantic dimension using the data for List 1 to develop the classification function for E+ vs E− semantic classes are given in the first two rows of the table. In this case, the classification function classified EPs to E+ and E− classes with 100% accuracy. The jackknifed crossvalidation success remained at 100% for both E+ and E− classes. The jackknifed procedure assesses the classification success when each case is left out of the development set and then classified. When the classification functions developed from List 1 data were applied to List 2 data, 80% of the EPs obtained to E+ words were correctly assigned to the E+ class and 80% of the E− EPs were assigned to the E− class. These percentages are to be contrasted with a chance level of 50%, since two classes at a time are considered in the unidimensional analyses.

The success in discriminating along the Potency semantic dimension (P+ vs P−) using List 1 data for development of the classification was also high. The percentages of P+ and P− EPs correctly classified for the data used in development and by the jackknifed crossvalidation were uniformly

TABLE 1

PERCENTAGES OF BRAIN RESPONSES (EPs) CORRECTLY CLASSIFIED IN SIX SEMANTIC CLASSES USING UNIDIMENSIONAL ANALYSES[a]

Semantic dimension	Pole	Development (List 1)	Jackknifed cross-validation (List 1)	Other-list cross-validation (List 2)
Evaluation	+	100	100	80
	−	100	100	80
Potency	+	100	100	100
	−	100	100	90
Activity	+	100	90	40
	−	90	80	50
Overall		98.3	95.0	73.3
		(List 2)	(List 2)	(List 1)
Evaluation	+	100	70	80
	−	80	70	50
Potency	+	100	90	100
	−	90	90	90
Activity	+	100	100	100
	−	100	90	20
Overall		95.0	85.0	73.3

[a] Each individual percentage is based on 10 EPs. The percentage correct expected by chance is 50%. χ^2 tests comparing each of the six overall success rates to chance yielded values from 12.2 to 54.2, each with 1 *df* ($ps << .001$).

100%. Furthermore, the classification functions developed from List 1 data were quite accurate in assigning List 2 data to P+ (100%) and P− (90%) semantic classes.

The accuracy of classifying EPs along the Activity semantic dimension was not as high as for the E and P semantic dimensions, although it was still quite respectable. The jackknifed crossvalidation success for List 1 was 90% for A+ and 80% for A− EPs. However, the A+ vs. A− classification function developed from List 1 data was not successful in classifying EPs from List 2: Correct assignments fell to chance levels (40 and 50%).

When the data for List 2 were used to develop the classification functions (the bottom half of Table 1), the results in general were quite similar to those obtained when the development was based on List 1 data. The differences are minor, with slightly lower percentages for the P semantic dimension and slightly higher percentages for the A dimension. The crossvalidation rates for the E dimension, however, were not significantly better than chance. In this case, one would do better using the classification rule developed with data on List 1.

Overall, the unidimensional analyses have an average apparent success of 97% and an average jackknifed crossvalidation success of 90%. It is to be noted that this success rate was obtained across subjects; the same classification functions were used for all 10 subjects. When the same classification functions were applied to the EP data obtained from the other word list, the overall success rate was 73%.

Multidimensional Analyses of Semantic Groups

Table 2 summarizes the results of classifying the brain responses into one of six semantic classes. The probability of correct classifications by chance is 1/6 (16.7% of the EPs). The classification success rates (column 1) for the EPs used to develop the functions were well above the chance level.

The success rates for the jackknifed crossvalidation (one-left-out procedure) also were above the chance level. Overall, they were 42% for List 1 and 43% for List 2 data, some 2.5 times better than chance.

When the classification functions developed from the data for one list were applied to the data for the other list, the overall success rates were both 40%. This is a stringent crossvalidation, since it assesses the ability to generalize not only to other EPs but also to a different list of words and to generalize across individual data of 10 different subjects.

A further test of the generalizability of the findings was made by applying the already computed classification functions unmodified to the EP data of a new subject. These data were collected and standardized in the same manner as for the 10 subjects in the development group. The component score coefficients obtained from the original 10-subject Principal Components Analysis were used to compute component scores for the new data. The classification functions obtained from the 10-subject Multiple

TABLE 2

PERCENTAGES OF BRAIN RESPONSES (EPs) CORRECTLY CLASSIFIED IN SIX SEMANTIC CLASSES USING MULTIDIMENSIONAL ANALYSES[a]

Semantic dimension	Pole	Development (List 1)	Jackknifed cross-validation (List 1)	Other-list cross-validation (List 2)
Evaluation	+	30	30	50
	−	80	50	50
Potency	+	40	40	30
	−	80	80	70
Activity	+	50	20	10
	−	60	30	30
Overall		56.7	41.7	40.0
		(List 2)	(List 2)	(List 1)
Evaluation	+	60	60	70
	−	70	60	50
Potency	+	30	20	20
	−	90	80	70
Activity	+	20	0	20
	−	50	40	10
Overall		53.3	43.3	40.0

[a] Each individual percentage is based on 10 EPs. The percentage correct expected by chance is 16.7%. χ^2 tests comparing each of the six overall success rates to chance yielded values from 21.9 to 66.3, each with *1* df ($ps << .001$).

Discriminant Analyses were applied to the new subject's EP component scores. The overall accuracy with which these EPs were correctly classified into the six semantic classes was 42%, essentially the same as the crossvalidation accuracy rates based on the EPs of the previous subjects.

DISCUSSION

Some EP research beginnings have been made at a variety of linguistic levels (Begleiter & Platz, 1969; Begleiter, Gross, & Kissin, 1967; Begleiter, Gross, Porjesz, & Kissin, 1969; Brown, Marsh, & Smith, 1973, 1976; Buchsbaum & Fedio, 1970; Chapman, 1974b, 1976; Chapman, Bragdon, Chapman, & McCrary, 1977; Feldman & Goldstein, 1967; Friedman, Simson, Ritter, & Rapin, 1975a, 1975b; Molfese, 1977; Molfese, Freeman, & Palermo, 1975; Molfese, Nunez, Seibert, & Ramanaiah, 1976; Shelburne, 1972, 1973; Teyler, Roemer, Harrison, & Thompson, 1973; Thatcher, 1977a, b; Wood, Goff, & Day, 1971; Wood, 1975; etc.). The only other EP studies known to be directed at the connotative meaning level of linguistics come from Begleiter and his associates. In general, the research seeking linguistic effects in EPs varies considerably in sophistication and conviction (Chapman, 1976). Alternative explanations of these EP effects in many cases are available in terms of sensory differences in the stimuli, different states of the subject, different cognitive functions, different sequence effects, and individual differences, etc.

Part of the evidence for the specificity of language effects in EPs depends on the *dimensionality* of the EP measures themselves. Since a prominent, late, positive-going component of the EP (variously called P300, P3, etc.) has been associated with the general relevance of stimuli to the subject's task (Chapman and Bragdon, 1964), it is relevant to ask whether the obtained linguistic effects in EPs are merely variations in this general EP component. The EP measures in the present study were obtained from a Principal Components Analysis which extracted 12 orthogonal components. One general EP component, such as P300 (Chapman, McCrary, Bragdon, & Chapman, in press), is not sufficient to account for the dimensionality of the present EP data. The classification functions used various combinations of a number (two to six) of the orthogonal EP components to distinguish the six semantic classes. Three of the components contributed to discriminations among semantic classes more strongly and consistently than others. One or more of these three contributed to each of the reported classifications. These were also the three components which collectively accounted for the highest proportion (39%) of the total variance in the EP data (all 12 accounted for 94%). All three of the components were principally correlated ($r > .32$) with the original EP measures at time points ranging from 245 to 510 msec, being maximally correlated ($r > .90$) around 295, 410, and 495 msec. In addition, depending upon the particular analysis, various numbers of other components contributed to the semantic discriminations. Only one of the 12 components was not used in any of the reported classifications. Thus, the semantic differences examined were not described by different amounts of any single EP component.

These findings suggest that internal representations of meaning can be assessed by analyzing electrical brain responses. Since the semantic classes were presented randomly, the obtained differences cannot be attributed to any prestimulus variables, e.g., expectancy, arousal, and attention, etc. Since the subject's task (perceive and say word) was constant, the obtained differences do not relate to general poststimulus variables, e.g., differential information processing, response preparation, and uncertainty resolution, etc. It is not likely that the EP differences are related to different muscle activity since (i) the words were spoken after the 510-msec EP interval and (ii) many different words constituted the stimuli for each semantic class. Analyses of the EOG data show that eye movements do not explain the EP effects. Since many different words were the stimuli for each semantic class, the number of luminous dots and alphabet distributions were similar across semantic classes, and the EP results were generalized across two such lists of words, it does not seem likely that the results are due to the physical differences in the visual stimuli. The same aspect of the experimental design guards against interpretations based on surface linguistic features. Finally, distinguishing six semantic classes indicates a degree of specificity which generally taxes interpretations in terms of variables other than connotative meaning.

The finding that the EP effects related to connotative meaning hold for all of the subjects suggests that the physiological representation of meaning may be similar in different individuals. Further substantiation of this finding would parallel, at the physiological level, the universality of the Osgood dimensions found by semantic differential ratings.

REFERENCES

Begleiter, H., Gross, M. M., & Kissin, B. 1967. Evoked cortical responses to affective visual stimuli. *Psychophysiology*, **3**, 336–344.

Begleiter, H., Gross, M. M., Porjesz, B., & Kissin, B. 1969. The effects of awareness on cortical evoked potentials to conditional affective stimuli. *Psychophysiology*, **5**, 517–529.

Begleiter, H., & Platz, H. 1969. Cortical evoked potentials to semantic stimuli. *Psychophysiology*, **6**, 91–100.

Brown, W. S., Marsh, J. T., & Smith, J. C. 1973. Contextual meaning effects on speech-evoked potentials. *Behavioral Biology*, **9**, 755–761.

Brown, W. S., Marsh, J. T., & Smith, J. C. 1976. Evoked potential waveform differences produced by the perception of different meanings of an ambiguous phrase. *Electroencephalography and Clinical Neurophysiology*, **41**, 113–123.

Buchsbaum, M., & Fedio, P. 1970. Hemispheric differences in evoked potentials to verbal and nonverbal stimuli in the left and right visual fields. *Physiology and Behavior*, **5**, 207–210.

Callaway, E. 1975. *Brain electrical potentials and individual psychological differences*. New York: Grune & Stratton.

Chapman, R. M. 1974. Latent components of average evoked brain responses functionally related to information processing. In *International symposium on cerebral evoked potentials in man, pre-circulated abstracts*. Brussels: Presses Universitaires de Bruxelles. Pp. 38–42. (a)

Chapman, R. M. 1974. Semantic meaning of words and average evoked potentials. In *International symposium on cerebral evoked potentials in man, pre-circulated abstracts*. Brussels: Presses Universitaires de Bruxelles. Pp. 43–45. (b)

Chapman, R. M. (Chair). 1976. *ERPS and language*. Transcript of the panel at the Fourth International Congress on Event Related Slow Potentials of the Brain (EPIC IV), David A. Otto, Program Chair., Hendersonville, N.C.

Chapman, R. M., & Bragdon, H. R. 1964. Evoked responses to numerical and non-numerical visual stimuli while problem solving. *Nature (London)*, **203**, 1155–1157.

Chapman, R. M., Bragdon, H. R., Chapman, J. A., & McCrary, J. W. 1977. Semantic meaning of words and average evoked potentials. In J. E. Desmedt (Ed.), *Progress in clinical neurophysiology*. Vol. 3: *Language and hemispheric specialization in man: Cerebral event-related potentials*. Basel: Karger. Pp. 36–47.

Chapman, R. M., McCrary, J. W., Bragdon, H. R., & Chapman, J. A. In press. Latent components of event-related potentials functionally related to information processing. In J. E. Desmedt (Ed.), *Progress in clinical neurophysiology*. Vol. 6: *Cognitive components in cerebral event-related potentials and selective attention*. Basal: Karger.

Dixon, W. J. (Ed.) 1975. *BMDP biomedical computer programs*. Berkeley: University of California Press.

Feldman, R. M., & Goldstein, R. 1967. Averaged evoked responses to synthetic syntax sentences. *Journal of Speech and Hearing Research*, **10**, 689–696.

Friedman, D., Simson, R., Ritter, W., & Rapin, I. 1975. Cortical evoked potentials elicited by real speech words and human sounds. *Electroencephalography and Clinical Neurophysiology*, **38**, 13–19. (a)

Friedman, D., Simson, R., Ritter, W., & Rapin, I. 1975. The late positive component (P300) and information processing in sentences. *Electroencephalography and Clinical Neurophysiology*, **38**, 255–262. (b)

Heise, D. R. 1971. *Evaluation, potency, and activity scores for 1551 words: A merging of three published lists*. Chapel Hill, N.C.: University of North Carolina, Department of Sociology.

Kaiser, H. F. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187–200.

Miron, M. S., & Osgood, C. E. 1966. Language behavior: The multivariate structure of qualification. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand–McNally. Pp. 790–819.

Molfese, D. L. 1977. The ontogeny of cerebral asymmetry in man: Auditory evoked potentials to linguistic and non-linguistic stimuli. In J. E. Desmedt (Ed.), *Progress in clinical neurophysiology*. Vol. 3: *Language and hemispheric specialization in man: Cerebral event-related potentials*. Basel: Karger. Pp. 188–204.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. 1975. The ontogeny of brain lateralization for speech and nonspeech stimuli. *Brain and Language*, **2**, 356–368.

Molfese, D. L., Nunez, V., Seibert, S., & Ramanaiah, N. V. 1976. Cerebral asymmetry: Changes in factors affecting its development. In S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds.), *Origins and evolution of language and speech. Annals of the New York Academy of Sciences*, **280**, 821–833.

Osgood, C. E. 1971. Exploration in semantic space: A personal diary. *Journal of Social Issues*, **27**, 5–63.

Osgood, C. E., May, W. H., & Miron, M. S. 1975. *Cross-cultural universals of affective meaning*. Urbana, Ill.: University of Illinois Press.

Regan, D. 1972. *Evoked potentials in psychology, sensory physiology, and clinical medicine*. New York: Wiley-Interscience.

Rummel, R. J. 1970. *Applied factor analysis*. Evanston, Ill.: Northwestern University Press. Pp. 246–247.

Shelburne, S. A., Jr. 1972. Visual evoked responses to word and nonsense syllable stimuli. *Electroencephalography and Clinical Neurophysiology*, **32**, 17–25.

Shelburne, S. A., Jr. 1973. Visual evoked responses to language stimuli in normal children. *Electroencephalography and Clinical Neurophysiology*, **34**, 135–143.

Teyler, T. J., Roemer, R. A., Harrison, T. F., & Thompson, R. F. 1973. Human scalp-recorded evoked-potential correlates of linguistic stimuli. *Bulletin of the Psychonomic Society*, **1**, 333–334.

Thatcher, R. W. 1977. Evoked potential correlates of delayed letter matching. *Behavioral Biology* **19**, 1–23. (a)

Thatcher, R. W. 1977. Evoked potential correlates of hemispheric lateralization during semantic information processing. In S. Harnad et al. (Eds.), *Lateralization in the nervous system*. New York: Academic Press. Pp. 429–448. (b)

Wood, C. C. 1975. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, **104**, 3–20.

Wood, C. C., Goff, W. R., & Day, R. S. 1971. Auditory evoked potentials during speech perception. *Science*, **173**, 1248–1251.

Short-Term Memory: The "Storage" Component of Human Brain Responses Predicts Recall

Robert M. Chapman, John W. McCrary, John A. Chapman
Center for Visual Science, University of Rochester
Rochester, New York 14627

Accepted for publication:
Science, in press, 1978

Abstract

An Evoked Potential component with a post-stimulus peak about 250 msec. is related to storage of information in short-term memory. This Storage Component was found in an experiment investigating brain potentials in relation to a number/letter information processing task. In replications of this experiment at three different light intensity levels spaced 1.0 log unit apart, essentially the same component waveform and pattern of component scores were found. The memory storage interpretation was confirmed in a behavioral experiment which probed short-term memory. Recall was predicted by the magnitude of the Storage Component.

A critical ingredient in most human information processing is the short-term storage of incoming information so that it can be integrated with other incoming information. Various kinds of memory processes with different properties have been proposed as retainers of information for various lengths of time [1], e.g. sensory register, short-term store, long-term store. Electrophysiological and behavioral evidence is presented here for a neural process which is related to storage in short-term memory. A latent component of electrically recorded brain responses (Evoked Potentials, EPs) with a post-stimulus peak about 250 msec. was found to be related to storage of stimulus information for later use in number and letter comparison tasks.

The Storage Component of the EP was discovered and tentatively interpreted as being associated with information storage in an experiment in which the latent components and component scores of the brain responses from 12 subjects were obtained by a Varimaxed Principal Components Analysis [2]. The Storage Component, which is the focus of this report, was one of eight orthogonal EP components obtained from that analysis.

The generality of the Storage Component and its independence of the physical characteristics of the stimuli were tested in further experiments [3]. The same stimuli and procedures were used except that the intensity of all stimuli within a run of 102 trials was ten times higher, the same, or one-tenth as high as in the original experiment [4].

Two numbers and two letters were flashed individually in random order at intervals of 3/4 sec. preceded and followed by a blank flash. The subject's task was to compare numerically the two numbers on number-relevant runs, the letters being irrelevant to the task. On the other half of the runs, the numbers were irrelevant and the task was to compare alphabetically the two letters [5].

The stimulus processing demanded by the task depended on a number of factors, including whether: (i) number or letter stimuli were task relevant, (ii) the number or letter class of stimulus could be anticipated, and (iii) the character was the first or second relevant stimulus of the pair to be compared. For the first relevant stimulus in each trial, the information had to be stored by the subject until the second relevant stimulus occurred, after which the comparison could be made.

While the subject was performing the letter or number comparison tasks, electrical brain activivity (EEG) was recorded from scalp electrodes [6].

By averaging the brain activity evoked by stimuli for similar conditions, averaged Evoked Potentials (EPs) were obtained for 16 conditions: relevant and irrelevant numbers and letters at four intra-trial positions. From trial to trial the first number (or letter) stimulus occurred in intra-trial position 1, 2 or 3, while the second number (or letter) stimulus occurred in intra-trial positions 2, 3, or 4. To simplify interpretations certain EEG data were discarded, so the EPs for intra-trial positions 1 and 2 were based only on the first number and letter stimuli presented within each trial, while the EPs for intra-trial positions 3 and 4 were based only on the second number and letter stimuli presented within each trial. For each of the three intensity levels, EPs were collected in the same manner and each of the three sets of data was analyzed separately [7]. Latent components and component scores of each of the data matrices were computed using varimaxed Principal Components Analysis [8].

In all three data sets, an EP component emerged which was strikingly similar to the Storage Component previously found [2] with regard to both waveform and relative magnitude for the 16 conditions (Fig. 1A, 1B, 1C). The waveforms of these components are very similar in all four sets of data, reaching their maximum about 250 msec. after the stimulus. The coefficients of factorial similarity among the waveforms from the four data sets were high, ranging between 0.85 and 0.99. For the previous experiment the maximum was at 250 msec.; for the new data the maxima are at 250 msec., 250 msec., and 270 msec. for high, mid, and low light intensities, respectively.

The Storage Component tends to be positive for stimuli whose information needs to be stored by the subject. Thus, the magnitude of the Storage Component was more positive for the first of the two relevant stimuli presented on each trial (intra-trial positions 1 or 2) than for the second relevant stimulus (intra-trial positions 3 or 4). The Storage Component was also relatively positive for the irrelevant stimuli when they occurred in intra-trial position 1. Extending the storage interpretation to this result leads to the hypothesis that an irrelevant stimulus in position 1 is stored in memory, whereas irrelevant stimuli in positions 2, 3, and 4 are not. This may be related to short-term memory having a limited capacity and storage of irrelevant information interfering with processing relevant information. The difference in the Storage Component scores for relevant and irrelevant stimuli in intra-trial position 2 is evidence that this component is not related simply to an order effect. Nor is this EP component related to amount of processing which is presumably greatest for the comparison operations following the second relevant stimulus, next most for the storage operations associated with the first relevant stimulus, and least for the irrelevant stimuli. Nor does this EP component reflect a general relevant-irrelevant distinction [9]. The Storage Component did not consistently distinguish between number and letter processing, or between number and letter stimuli. The simplest and most direct interpretation is that this EP component is related to the storage of information in the subject's short-term memory. More specifically the component may reflect the process of reading information out of a sensory register into short-term memory. Not only were the Storage Component scores related to memory storage conditions, but also the timing of this EP component (waveforms in Fig. 1 insets) is appropriate for information storage. The maximum of the Storage Component was at 250 msec. This is an appropriate time for storing information needed later since the literature suggests that the sensory register (icon) is fading about that time [10].

The results demonstrate the robustness of the Storage Component in the face of large differences in the physical parameters of the stimuli. That the Storage Component represents neural activity in the stimulus-response sequence that occurs later than the simple processing of sensory input is supported by two findings: (i) its independence of whether the stimuli are numbers or letters and (ii) that changes in stimulus intensity which are sufficient to alter markedly the overall EP have only a small effect on the Storage Component. Further, that the Storage Component occurs after the simple processing of sensory input, including recognition of the informational content of the stimulus, is indicated by the differences in response to identical physical stimuli when they play different roles in the information processing task. For example, in Fig. 1 compare the component scores to relevant and irrelevant stimuli in intra-trial position 2 and compare the component scores to relevant stimuli in positions 1 or 2 with those in positions 3 or 4. Along the time continuum, the Storage Component precedes both the behavioral response and the comparison operations which cannot occur until intra-trial positions 3 or 4. The Storage Component (maximum about 250 msec.) occurs before an EP component related to alphabetic comparison (maximum at about 350 msec.) [2].

Our tentative interpretation that this EP component is related to storage was based on considering the differential scores for the 1st and 2nd relevant stimuli within each trial. However, finding high Storage Component scores for irrelevant stimuli in position 1 required ad hoc interpretations in order to maintain the storage identification. Therefore, we felt that it was important to check more directly the storage interpretation by a behavioral experiment designed to assess storage in short-term memory.

The behavioral experiment used a memory probe technique to test the subjects' recall of individual stimuli for each of the 16 conditions in the electrophysiological experiments. Experimental sessions and data collection were conducted by experimenters not involved in the previous experiments and not aware of the hypothesis being tested. The experimental procedure was the same as for the collection of brain responses with the addition of occasional memory probes. The primary task on each trial was to compare the two numbers on one run of 102 trials and to compare the two letters on a second run [11]. Within each run of 102 trials, eight randomly located memory probes were selected to test recall of a letter and a number in each of the four intra-trial positions. Without prior warning of when probes would occur, blank flashes were delivered 3/4 and 1 1/2 sec. after the probed stimulus and the subject was asked what the last character was. These blank flashes were used to mask the probed stimulus and to delay the recall report in order to reduce the effects of very short-term sensory register. From each subject, one such recall probe was obtained for each of the 16 conditions (8 probes each in a number-relevant and a letter-relevant run). The percent correct recalls from 52 subjects (29 female and 23 male college students) are given in z-score units in Fig. 1D [12]. The pattern of correct recalls is strikingly similar to the pattern of Storage Component scores in Fig. 1A, 1B, 1C with better memory for relevant stimuli in intra-trial positions 1 and 2 and irrelevant stimuli in position 1. The six correlations among the four patterns of 16 means each in Fig. 1 ranged from .71 to .97. Three interesting features of the data are common to both the Storage Component of the brain responses and the subjects' short-term memory: (1) the first relevant stimulus on a trial (intra-trial position 1 or 2) gave higher scores than did the second relevant stimulus (positions 3 or 4); (2) the scores were high for both relevant and irrelevant stimuli in intra-trial position 1; and (3) in position 2, the scores were higher for relevant stimuli than irrelevant stimuli. The recall performance is plotted as a function of mean Storage Component score (averaged over the three intensity levels) in Fig. 2. Thus, the storage interpretation was confirmed by predicting recall performance on the basis of the Storage Component of brain responses ($r = .77$). The accuracy of this prediction is impressive considering that behavioral recall is not solely a function of storage but is generally considered to be greatly influenced by other factors including retrieval mechanisms.

One of the reasons that the Storage Component has not been found in other EP research is that it may be partially masked by a positive peak in the EP which often occurs slightly before 250 msec. Hence, measurement based on peaks of the average EP may miss the latent Storage Component that was derived by Principal Component Analyses which assess the relationships among all the time points and decompose EPs into independent sources of variation. Now that the Storage Component has been described and its waveform is known, it may be measured by computing component scores directly in other EP studies without doing a complete Principal Component Analysis.

Regardless of the reasons why stimulus information sometimes is not stored in an individual's memory, it would be extremely useful from theoretical, experimental, educational and clinical standpoints to be able to determine whether or not stimulus information is being stored in memory. If further research sustains the interpretation that the Storage Component of Evoked Potentials reflects the process of storing information in short-term memory, then this brain response component may be used to assess storage per se, uncontaminated by retrieval mechanisms. The Storage Component of Evoked Potentials holds promise of serving this practical function as well as providing an entry to understanding the neural processes related to memory.

References and Notes

1. E.g., R. C. Atkinson and R. M. Shiffrin, in The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 2, K. W. Spence and J. T. Spence, Eds. (Academic Press, New York, 1968), pp. 89-195.

2. R. M. Chapman, in International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts, (Presses Universitaires de Bruxelles, Brussels, 1974), pp. 38-42. R. M. Chapman, J. W. McCrary, H. R. Bragdon, and J. A. Chapman, in Progress in Clinical Neurophysiology. Vol. 6: Cognitive Components in Cerebral Event-Related Potentials and Selective Attention, J. E. Desmedt, Ed. (Karger, Basel, in press.) The analysis model assumes that the EPs are linear, weighted sums of a number of independent components, each with a fixed time course. The time courses of the components and the weights (component scores) are derived from the data by intercorrelating the responses at every time point and factoring the resulting correlation matrix of 102 time points with unities retained in the diagonal. Loosely speaking, time points which vary together are identified as belonging to the same component; any time point may partly belong to more than one component. The number of components was limited by the 'eigenvalue = one' criterion, and rotation was performed using normalized varimax criterion [H. Kaiser, Psychometrika **23**, 187-200 (1958)].

3. Preliminary reports in R. M. Chapman, Electroenceph. clin. Neurophysiol. **43**, 778 (1977) and R. M. Chapman, in Visual Psychophysics: Its Physiological Basis, J. C. Armington, J. Krauskopf and B. R. Wooten, Eds. (Academic Press, New York, 1978 in press), pp. 469-480.

4. Light intensity was controlled by interposing Wratten neutral

density filters. The middle light intensity was approximately 0.1 foot-lambert and was about 2.0 log units above threshold for character recognition.

5. By appropriately moving a two-way switch at the end of each trial, the subject indicated whether the first or second number was larger on number-relevant runs and similarly indicated the alphabetic order on letter-relevant runs. The numbers and letters were randomly selected (1-6, A-F) and the sequences of numbers and letters were randomized. Nearly every stimulus was processed appropriately by the subjects, with a performance accuracy of better than 99%. All stimuli were flashed at the same spatial location by a Bina-View display equipped with a Grass strobe. More details are in R. M. Chapman, in Psychophysiology of Thinking, F. J. McGuigan and R. Schoonover, Eds. (Academic Press, New York, 1973), pp. 69-108.

6. Data reported here were recorded monopolar from the midline central-parietal area (CPZ) to linked ear lobes. Frequency band-pass was .3 to 70 Hz; 102 samples at 5 msec. intervals were obtained beginning 30 msec. before each stimulus.

7. Each EP consisted of amplitude measurements in microvolts at 102 time points spanning 510 msec. The data were collected from one subject over a series of ten sessions. The data set at each intensity contained 160 EPs (16 conditions by 10 replications), each based on averaging the EEG to 34-51 stimuli (depending on the intra-trial list position).

8. Same method as in [2]. BMDP4M Program in W. J. Dixon, Ed., BMDP Biomedical Computer Programs (University of California Press, Berkeley, 1975.)

9. The Storage Component was orthogonal to a number of other uncorrelated EP components, including P300 and CNV-resolution,

which were also found in the analyses. Thus, this Storage Component is not just a variant of well-known EP components, but rather is orthogonal to them in a formal analysis. Also, this Storage Component was not correlated with eye movements (EOG) or with amount of alpha EEG.

10. G. Sperling, Psych. Monogr. 74, No. 11 (Whole No. 498) (1960); A. O. Dick, Perception & Psychophysics 16, 575 (1974).

11. The order of number-relevant and letter-relevant runs was reversed for half the subjects. Practice trials on the primary task were given until ten consecutive trials were correct. The average performance on the primary task was 97.5% correct. The light intensity was approximately 2.8 log units above threshold for character recognition.

12. Percent correct were converted to probit scores (z-score units), e.g. 50% and 98% become 0.0 and +2.05 probit scores.

13. Henry R. Bragdon collected the data in the evoked potential experiment. Debra Gershowitz and Janice K. Martin collected the data in the behavioral memory-probe experiment. Supported in part by PHS grants EY01593 and EY01319 and ONR contracts N00014-77-C-0037 and CNA SUB N00014-76-C-0001.

Figure Legends:

Fig. 1. Storage Component of brain responses and recall for experimental conditions in which short-term memory demands vary. A letter (L) or number (#) was flashed in each of the intra-trial positions (spaced 3/4 sec.). The task (compare letters or numbers) required short-term memory for relevant stimuli (circled) in positions 1 and 2. Brain response and behavioral measures are in z-score units. Note the similarity of the pattern of the data in all four panels. (A), (B), (C): Evoked Potential (EP) Storage Components were obtained from separate varimaxed Principal Components Analyses on EPs obtained with stimuli at three light intensities spaced 1.0 log unit apart. Insets show the Storage Component waveforms scaled appropriately for relevant numbers in position 1; the fundamental time course of the component (rotated factor loadings multiplied at each of 102 time points by standard deviations) was multiplied by the mean Storage Component score for that condition; the 480 msec. calibration bar begins at the stimulus flash. Storage Component waveforms peaked at 250, 250, 270 msec. for high, mid, and low intensities, respectively. EPs obtained from scalp electrode at CPZ (central-parietal midline) referred to linked earlobes. (D): Mean recall by 52 subjects using an occasional, random memory probe ("What was the last character?") while performing the primary task of comparing numbers or letters. Percent correct recall converted to equivalent z-score (probits).

Fig. 2. Behavioral recall as a function of brain response Storage Component score. Pearson correlation coefficient is .77; linear regression line shown. Mean EP component score is average of

Storage Component scores found at the three stimulus intensities (Fig. 1A, 1B, 1C). Mean recall from 52 subjects obtained by memory probe (Fig. 1D). Experimental conditions: letters (L) or numbers (#); intra-trial positions (1, 2, 3, or 4); relevant to primary task (circled) or irrelevant (not circled).

Fig. 1

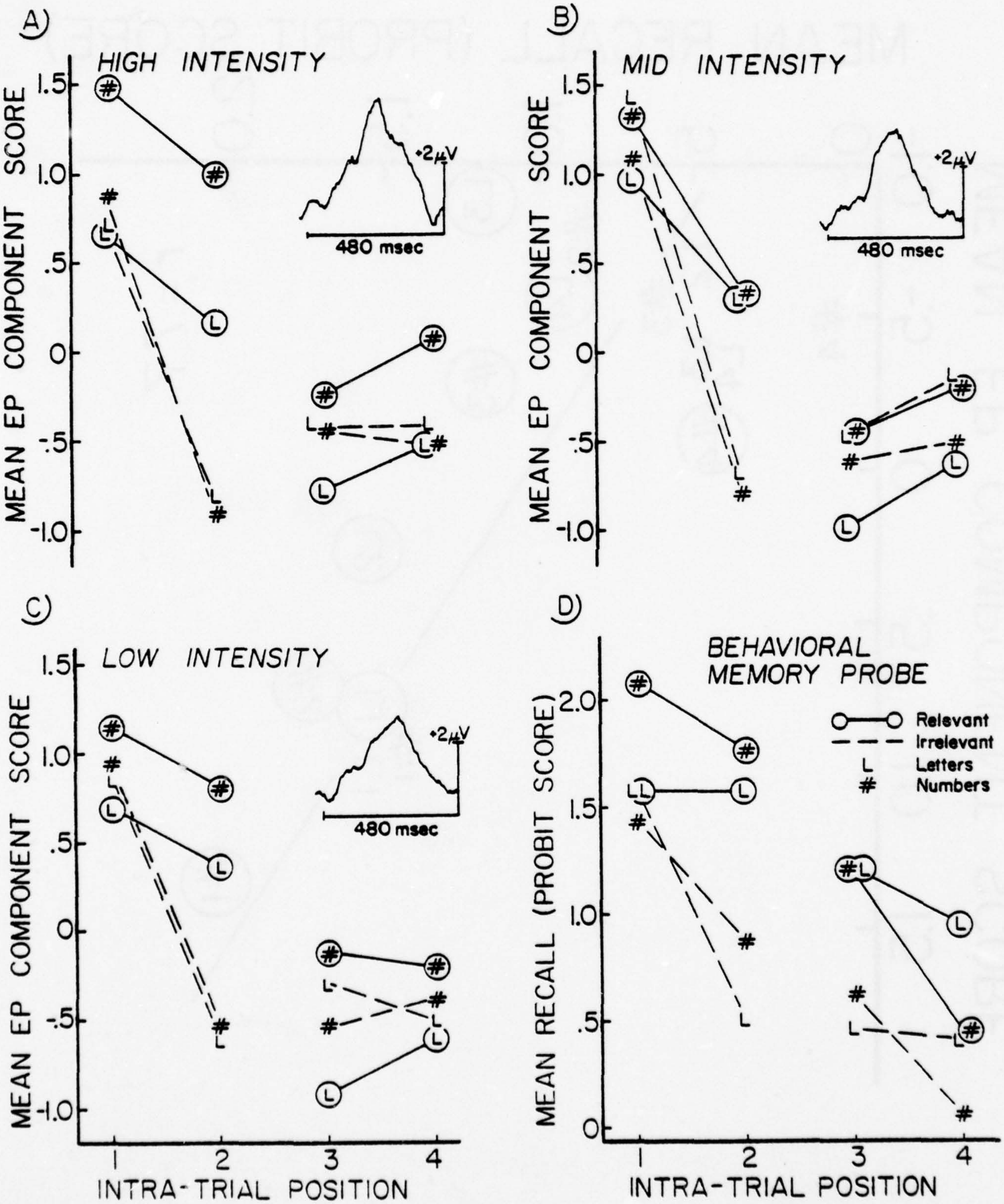
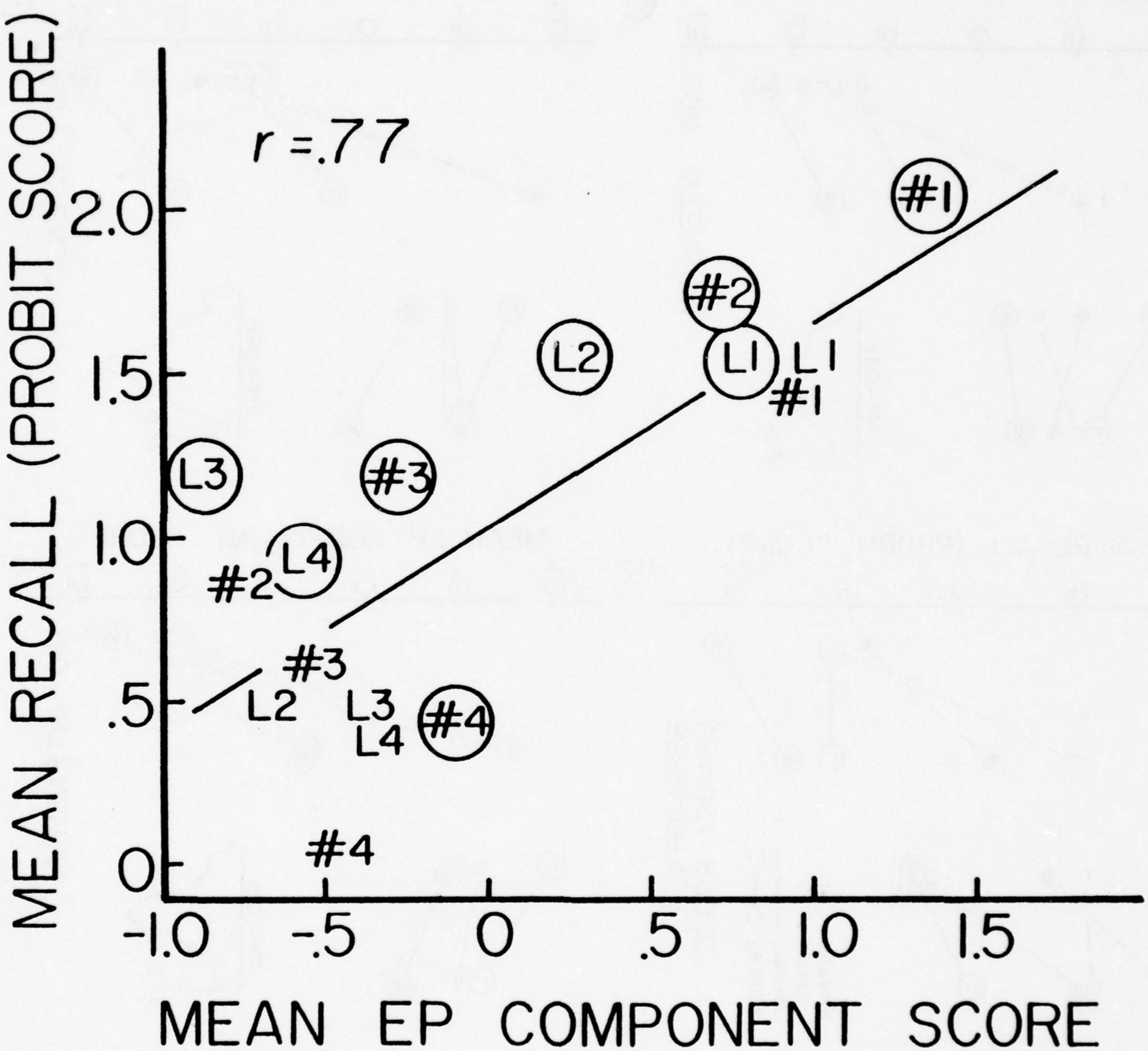

A)
HIGH INTENSITY
B)
MID INTENSITY
C)
LOW INTENSITY
D)
BEHAVIORAL MEMORY PROBE
MEAN EP COMPONENT SCORE
MEAN RECALL (PROBIT SCORE)
+2μV
480 msec
Relevant
Irrelevant
L Letters
Numbers
INTRA-TRIAL POSITION

Fig. 2


r =.77
MEAN RECALL (PROBIT SCORE)
2.0
1.5
1.0
.5
0
-1.0
-.5
0
.5
1.0
1.5
MEAN EP COMPONENT SCORE
#1
#2
L1
L1
#1
L2
L3
#3
L4
#2
#3
L2
L3
L4
#4
#4

In E. Callaway and D. Lehmann (Eds.), Event Related Potentials in Man: Applications and Problems. New York: Plenum Press, in press.

(Based on Conference in Konstanz, Germany, August, 1978)

HEMISPHERIC DIFFERENCES IN EVOKED POTENTIALS TO RELEVANT AND IRRELEVANT VISUAL STIMULI

Robert M. Chapman and John W. McCrary

University of Rochester

Rochester, New York 14627

Since the early days of averaging Evoked Potentials (EPs) in man, the importance of cognitive variables, as well as stimulus variables, has been recognized (e.g., Chapman and Bragdon, 1964). Using an experimental design which involves processing number and letter stimuli, we have been studying EP effects related to a variety of cognitive operations (Chapman, 1965; 1966; 1969a; 1969b; 1973; 1974a; 1974b; 1977; in press; Chapman, McCrary, Bragdon, and Chapman, in press; Chapman, McCrary, and Chapman, in press). Most of our analyses have been for the CPZ scalp location (recorded monopolar on the midline 1/3 of the distance from Cz to Pz; reference was linked ear lobes). It is of interest to study the cognitive effects at other sites, with a particular focus on the question of hemispheric differences and parietal-occipital differences.

A more complete description of the experimental design and discussion of interpretations for the present chapter is given in Chapman (1973). In that paper results are given for 12 subjects for midline electrodes located over the central-parietal (CPZ) and the occipital area (Oz), as well as control data for EOG and alpha EEG. The present experiment provides comparable data for 8 subjects for laterally located electrodes over parietal (P3 and P4) and occipital (O1 and O2) areas, and permits an evaluation of hemispheric differences in the information processing tasks. In general, comparable information-processing effects were found in both experiments. The evaluation of location differences was facilitated by the addition of control EPs to blank flashes and the use of additional analysis procedures, featuring Discriminant Analyses.

Earlier work on hemispheric specialization has been critically reviewed by Donchin, McCarthy & Kutas (1977). A caveat should be noted in considering hemispheric differences, or any brain localization effects, from EP data. EP effects localized at some scalp site do not necessarily mean that the adjacent brain region is responsible for those processes. Because the measure is a voltage difference in an electrical field of a conducting medium, the orientation of the source as well as its distance are important. Far field effects have been demonstrated for early auditory potentials (Jewett et al., 1970). The importance of source orientation is illustrated by scalp localizations opposite to brain hemisphere in visual field studies (Halliday et al., 1977). Incidentally, the same problems exist for electrical recording within brain structures as for scalp recording. Given this caveat, the spatial localization interpretations given in this chapter, strictly speaking, refer to particular scalp sites (with ear reference) and should be extended to brain localization with great caution.

Another problem relates to the assumption that larger EP amplitudes signify more processing. We suggest a method of analysis here which avoids this assumption, at least in its usual simplistic form. The method is based on Discriminant Analyses which focus on variations of EP measures which maximally discriminate particular conditions. This approach does not rely on sheer amplitude, but rather seeks combinations of amplitudes, large or small, which most systematically covary with particular sets of experimental conditions.

EXPERIMENTAL PROCEDURE

Two numbers and two letters were flashed individually in random order at intervals of 3/4 sec. preceded and followed by a blank flash. The subject's task was to compare numerically the two numbers on number-relevant runs, the letters being irrelevant to the task. On the other half of the runs, the numbers were irrelevant and the task was to compare alphabetically the two letters. By appropriately moving a momentary two-way switch at the end of each trial, the subject indicated whether the first or second number was larger on number-relevant runs and similarly indicated the alphabetic order on letter-relevant runs. The subject had a 1.5 sec. time slot following the last flash in which to answer before the next trial started. Correct answers produced a tone; wrong answers produced a buzz. The numbers and letters were randomly selected (1-6, A-F) and the sequences of numbers and letters were randomized. Nearly every stimulus was processed appropriately by the subjects, with a performance accuracy of better than 99%. All stimuli were flashed at the same spatial location by a Bina-View display equipped with a Grass strobe (flash duration < 10 microsec.).

The stimulus processing demanded by the task depended on a number of factors, including whether: (i) number or letter stimuli were task relevant, (ii) the number or letter class of stimulus could be anticipated, and (iii) the character was the first or second relevant stimulus of the pair to be compared. For the first relevant stimulus in each trial, the information had to be stored by the subject until the second relevant stimulus occurred, after which the comparison could be made.

While the subject was performing the letter or number comparison tasks, electrical brain activivity (EEG) was recorded from scalp electrodes at P3, P4, O1, and O2 (referenced to linked ear lobes). Frequency band-pass was 0.3 to 70 Hz; 102 samples at 5 msec. intervals were obtained beginning 30 msec. before each stimulus. The data were collected from eight right-handed subjects (5 male, 3 female) over a series of six sessions each.

By averaging the brain activity evoked by stimuli for similar conditions, separate averaged Evoked Potentials (EPs) were obtained for 16 information-processing conditions: relevant and irrelevant numbers and letters at four intra-trial positions. From trial to trial the first number (or letter) stimulus occurred in intra-trial position 1, 2 or 3, while the second number (or letter) stimulus occurred in intra-trial positions 2, 3, or 4. To simplify interpretations certain EEG data were discarded, so the EPs for intra-trial positions 1 and 2 were based only on the first number and letter stimuli presented within each trial, while the EPs for intra-trial positions 3 and 4 were based only on the second number and letter stimuli presented within each trial.

Even the irrelevant stimuli in this experiment must be processed to a certain extent to determine that they are irrelevant. The subject cannot anticipate whether the stimulus will be a letter or a number, and hence relevant or irrelevant, except in intra-trial position 4. To provide a control with even less processing by subjects, runs were added in which only blank flashes occurred. The blank flashes were provided by the same Bina-View device and appeared as an illuminated rectangle. The trials for those runs had the same temporal structure as the letter-number trials: blank flashes at the 4 intra-trial positions, preceded and followed by a blank flash, all spaced 3/4-sec. apart.

Each run contained 102 trials, each with four intra-trial positions. Each subject was given 10 number-relevant, 10 letter-relevant, and 4 blank runs spaced over a number of sessions. Averaging across all runs, the EPs for each subject were based on the EEG responses to 272 to 510 stimuli. This yielded 20 EPs for each subject: relevant and irrelevant numbers and letters and blanks for each of the 4 intra-trial positions. For each electrode, the data set consisted of 160 EPs (20 x 8 subjects).

EP Measures

The EPs were measured in the manner described in Chapman (1973) in order to facilitate comparison with the midline results reported there. For each EP, five measures were obtained: mean amplitude over 480 msec., amplitude at 0 msec., and amplitude at 105 msec., 225 msec., and 315 msec. The most global measure was mean amplitude over 480 msec. relative to a baseline obtained at 0 msec. (time of stimulus; the baseline was the average of 4 time points before and 3 after the stimulus). The amplitude at 105 msec., 225 msec., and 315 msec. were similarly measured relative to the same baseline at 0 msec. These measures index the amplitude at specified points within the EPs without the necessity

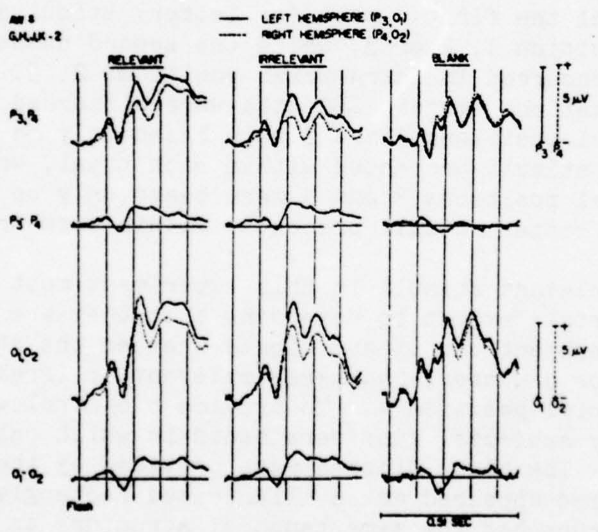


Fig. 1. Sample Evoked Potentials from one subject. Monopolar recording from left and right parietal (P3,P4) and occipital (O1,O2) scalp locations (referenced to linked ear lobes). Vertical lines 100 msec. apart.

of identifying particular peaks. The amplitude at 0 msec. was measured relative to an arbitrary voltage level across the entire trial of 4 intra-trial positions. The amplitude at 0 msec. indexes CNV activity.

RESULTS

Fig. 1 illustrates some of the EPs for one of the subjects. For this figure, the EPs were averaged across numbers and letters

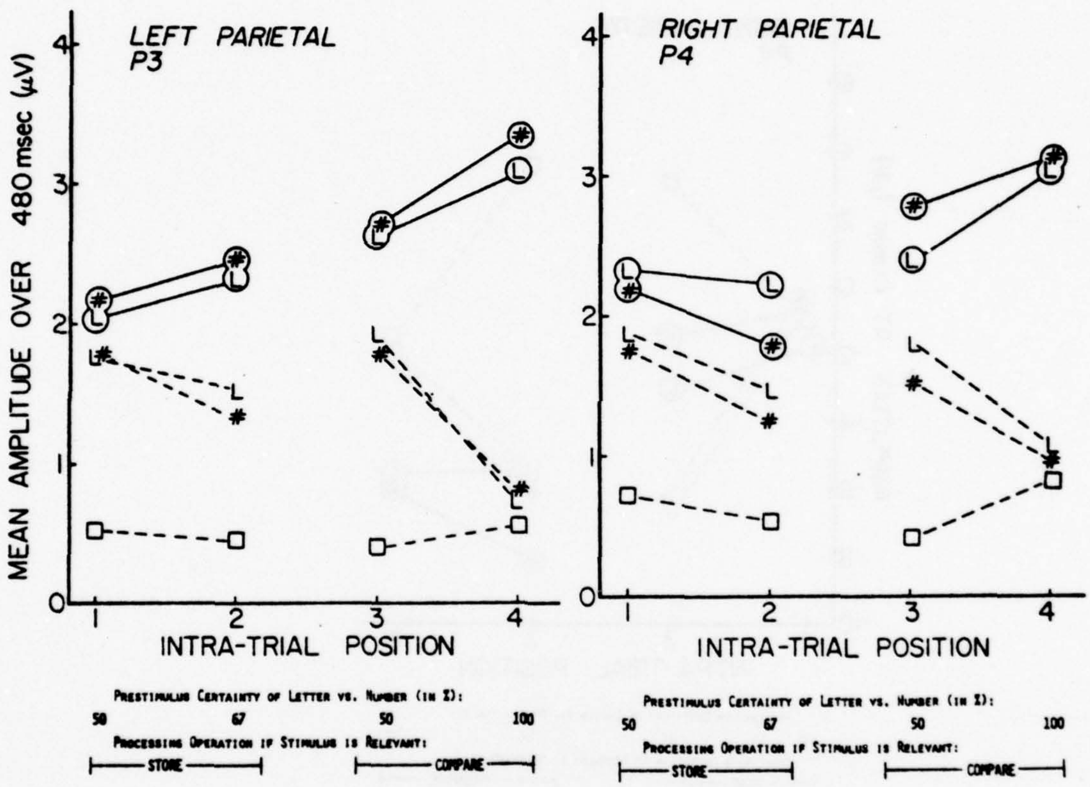


Fig. 2. Mean amplitude over 480 msec. from left and right parietal electrodes for 20 experimental conditions with varying information processing demands. Number (#), letter (L), and blank (box) visual stimuli. Relevant (circled symbols and solid lines) and irrelevant (not (circled symbols and dashed lines). Information processing characteristics associated with intra-trial positions are summarized below the abscissa. Data are means from 8 subjects.

and intra-trial positions, in order to illustrate the hemispheric differences for relevant, irrelevant, and blank stimuli. In this case, the EPs from the left are larger than those from the right, and this hemispheric difference is greater for relevant and irrelevant stimuli than for blank stimuli. Drawing conclusions from the data of one subject may be misleading. To assess those effects which have more generality, the data for all eight subjects have been examined as a set.

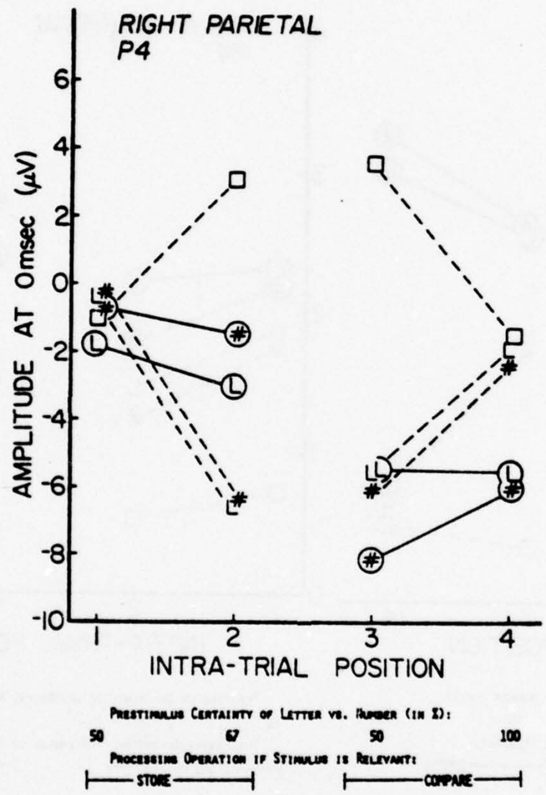


Fig. 3. Amplitude at 0 msec. relative to an arbitrary voltage level which was the same for all responses. This measure indexes CNV. Other specification as for Fig. 2.

EP Measures for Experimental Conditions

The results for mean amplitude over 480 msec. are quite similar from left and right electrodes (P3 and P4 shown in Fig. 2) and are similar to those previously obtained from midline electrodes at CPZ and Oz (Chapman, 1973, Figs. 3.6 & 3.7). The most striking result is the difference between relevant and irrelevant stimuli, regardless of whether numbers or letters were involved. There is also an interaction between relevance and intra-trial position. In addition, the EPs to the blank flashes are considerably smaller than the responses to the number and letter stimuli. However, the EPs to the irrelevant numbers and letters in intra-trial position 4, where there is 100% prestimulus certainty of stimulus class, approach the low amplitudes obtained to the blank flash controls.

Although there appear to be differences between the results for P3 and P4, the similarities dominate comparisons. The results for O1 and O2 (not shown) are also quite similar.

The amplitude at 0 msec. showed a different pattern of relations to the experimental conditions (Fig. 3) which was similar to midline data previously reported (Chapman, 1973, Fig. 3.12). There were essentially no differences between relevant and irrelevant conditions at intra-trial positions 1 and 3. At these positions, there was a 50-50 chance of a letter or number occurring and therefore a 50-50 chance of the stimulus being relevant or irrelevant. However, the prestimulus certainty of a letter or number occurring in intra-trial positions 2 and 4 was biased (67% and 100%, respectively). At positions 2 and 4 there was a difference in amplitude at the time of the stimulus for relevant and irrelevant stimuli. At intra-trial position 4, where there was 100% certainty prior to the presentation of the stimulus, the amplitude at 0 msec. was more negative when the stimulus was to be relevant than when it was to be irrelevant. This result is in agreement with the CNV literature, in which a negative potential is found in anticipation of an "imperative" (relevant) stimulus. The results at the other electrode sites for this measure were similar (P3, O1, O2 not shown). Hemispheric differences were not prominent.

The other EP measures, amplitudes at 105 msec., 225 msec. and 315 msec., showed major effects similar to those previously reported for midline electrodes (Chapman, 1973). Hemispheric differences were not pronounced. The measure which showed the most pronounced hemispheric differences was the amplitude at 315 msec. (Fig. 4). The pattern of data at 315 msec. suggests there may be differential hemispheric and brain area representation of various information processing conditions. The most obvious of these is a differential interaction of stimulus relevance and intra-trial position with hemisphere ($F=8.08$, $df=3,21$, $p<.01$). The question remains whether there is more differential representation of information processing in one hemisphere than in the other.

Hemispheric Differences in Discriminant Analysis

One way to assess whether responses from one area or another are more involved in various functions is by the use of Discriminant Analyses. If measures of responses from two (or more) brain areas are used to discriminate two (or more) experimental conditions, which measures do the best job?

Do EPs from the left (P3, O1) or right (P4, O2) hemisphere do a better job in discriminating various number/letter information processing conditions (relevant and irrelevant numbers and letters at four intra-trial positions). In the first application of the technique to be described, there are 16 classes to be discriminated from each other. To perform this discrimination, there are available 5 measures from each of 4 electrodes (20 variates). The stepwise Discriminant Analysis (BMDP7M, Dixon, 1975) selects the measures in the order of their effectiveness in classifying each of the 128 responses into the 16 experimental conditions. The intercorrelations among the measures are taken into account. For the next measure to be added to the prediction equation, the stepwise procedure selects the measure which is most effective after the influence of the previously selected measures is taken into account. When the Discriminant Analysis is allowed access to all 20 measures, the single best measure in discriminating the 16 information processing conditions was the Mean Amplitude Over 480 msec. from P3 (left parietal area). Of the first seven measures, six were from the left hemisphere (P3 Mean Over 480 msec., P3 at 0 msec., P3 at 315 msec., O1 at 315 msec., P4 Mean Over 480 msec., O1 at 105 msec., P3 at 105 msec., in order of their selection). Since there were 16 conditions to be discriminated, chance was 1/16 or 6.25%. The development classification success using the first seven measures was 47.7%. A better index of the generality of the success rate is the jackknifed classification success which was 28.1% (Table I). The jackknifed procedure is a cross-validation technique which assesses the classification success when each case is left out of the development set and then classified. This success rate is significantly better than chance (Chi-square = 100.8, $df=1$, $p<<.001$).

Another assessment of hemispheric differences involves computing separate Discriminant Analyses with measures from each side alone and comparing the classification success rates. The results of this procedure also are given in Table I. When

TABLE I

Discrimination of Experimental Conditions Using EP Measures from Both Sides, Left Side, and Right Side.

Groups	Chance	Both Sides	Left Side	Right Side
Information Processing				
16: number or letter X relevant or irrelevant X 4 intra-trial positions	6.25%	28.1% (6L,1R)	28.1% (5P,3O)	20.3% (3P,3O)
Information Processing				
9: relevant or irrelevant X 4 intra-trial positions & blanks	12.0%	53.1% (5L,3R)	51.9% (3P,2O)	46.9% (5P,2O)
Relevance				
3: relevant, irrelevant, and blanks	36.0%	85.0% (5L,3R)	81.9% (2P,2O)	76.9% (5P,2O)
Stimuli, physical				
3: numbers, letters, and blanks	36.0%	70.6% (6L,2R)	71.2% (4P,4O)	63.1% (3P,3O)
Individual Subjects				
8: subject	12.5%	96.9% (2L,8R)	92.5% (5P,5O)	94.4% (5P,5O)

Entries are jackknifed classification success rates (maximum for 10 or less variates) from Stepwise Discriminant Analyses (BMDP7M). All were significantly better than chance. The values of Chi-square (1 df), corrected for discontinuity, ranged from 40.7 to 1033.7 ($p \ll .0001$). Below each percentage, the number of left and right variates (L & R) or number of parietal and occipital variates (P & O) used in the classification functions are given in parentheses. The response measures were standardized separately for each of the subjects before performing the Discriminant Analyses except for the individual subject's analyses. Each subject's data for each measure were transformed to z scores with mean equal to 0 and stan. dev. equal to 1. This procedure has been found useful in reducing the effect of individual differences upon subsequent analyses which focus on the effect of experimental conditions (Chapman, McCrary, Chapman & Bragdon, 1978). The general conclusions reached with the subject-standardized measures are the same as those obtained with the raw measures; the main differences are improved rates of classification success when irrelevant subject differences have been removed.

discriminating the 16 information processing conditions, the measures from the left side alone (P3, O1) achieved the same classification success as when measures from both left and right sides were available (28.1%). A lower classification success rate (20.3%) was obtained when measures from the right side alone (P4, O2) were used. These results indicate that measures from both left and right sides carry information about the information processing conditions, but that the left-side measures carry more such information than those from the right side. The fact that the left side alone does as well, or nearly as well, as when both sides could contribute to the classification equations indicates that the measures from the right side are largely redundant with those from the left side. The single most important variate of the ten available from each side was the Mean Amplitude Over 480 msec. from the parietal site (P3 for left side alone, P4 for right side alone).

Essentially the same pattern of results was obtained for additional groupings of the experimental conditions (Table I). In order to provide comparisons which included the blank control flashes, the information processing design was simplified by ignoring whether the stimuli were letters or numbers. When discriminating the blanks and the resulting 8 information processing conditions (relevant or irrelevant stimuli x 4 intra-trial positions), the single best measure was again found to be the Mean Amplitude Over 480 msec. from P3. The first four measures selected for inclusion in the discrimination were from the left hemisphere. The final set of variables selected included five from the left and three from the right and accurately classified (jackknifed) 53.1% of the cases. Restricting selection of variates to the left side reduced the classification accuracy only slightly. Selecting variates only from the right produced a somewhat larger reduction (Table I).

Various kinds of functions may be assessed in a similar manner by using appropriate classification groups. For example, the side more related to stimulus relevance, regardless of stimulus or intra-trial position, was assessed by Discriminant Analyses using three groups: relevant, irrelevant, and blanks (Table I). The results suggest that the left-side EPs carry more information concerning stimulus relevance (81.9%), but that right-side EPs also do a good job in discriminating relevance (76.9%).

Which side was more related to the different physical stimuli was assessed by discriminating three groups: numbers, letters, and blanks (regardless of relevance or intra-trial position). The results indicate that the variates from the left side are more related to differences among the visual stimuli (Table I). The single most important variate was the amplitude at 315 msec. from the left occipital area (O1).

It is possible to use this Discriminant Analysis technique to assess which is more related to individual differences. For this purpose the groups were the eight individual subjects. For these analyses the raw measures, before subject standardization, were used. Classification functions were computed which classified each EP case to one of the subjects, regardless of the experimental conditions (relevant and irrelevant numbers and letters, and blanks, in four intra-trial positions). When measures from both sides were available, 96.9% of the EP cases were correctly classified to the individual subject by discriminant functions using two left variates and eight right variates. Measures from the left side alone did not do as well as measures from the right side alone (92.5% and 94.4%, respectively). This evidence suggests that the right side is more closely related to individual differences.

In general, the results indicate that measures over both hemispheres do a reasonably good job of discriminating various experimental conditions and individuals. The classification accuracy is well above chance in every instance. When discriminating information processing characteristics, variates from the left hemisphere are consistently selected first and often for inclusion in the discriminant equations. Although the differences are not statistically reliable, accuracy is consistently reduced when only variates from the right hemisphere are used in the discrimination. This consistency suggests that measures from the left side are more related to various information processing distinctions than measures from the right side. Measures from the right side appear to be more related to individual differences.

CONCLUSIONS

In a number-letter information processing experiment, comparing laterally recorded EPs with each other, and comparing the lateral EPs with previously reported midline EPs, the similarities are more striking than the differences. However, rather subtle hemispheric differences which are reasonably consistent have been found. The assessment of these lateral effects was facilitated by the use of control stimuli (blank flashes) and by particular kinds of Multiple Discriminant Analyses. These have provided evidence that some kinds of processes are more strongly related to the left side, while other processes are not. Information processing, including stimulus differences, was more discriminated by EP measures from the left side. Individual differences were more related to the right side.

REFERENCES

Chapman, R.M. Evoked responses to relevant and irrelevant visual stimuli while problem solving. Proceedings of the American Psychological Assoc., 1965, 177-178.

Chapman, R.M. Human evoked responses to meaningful stimuli. XVII Internat'l Congress of Psychology, Moscow, 1966, 6, 53-59.

Chapman, R.M. Discussion on eye movements, CNV and AEP. In E. Donchin and D.B. Lindsley (Eds.) Averaged evoked potentials: Methods, results and evaluations. Washington, D.C., NASA SP-191, 1969a, 177-180.

Chapman, R.M. Definition and measurement of "psychological" independent variables in an average evoked potential experiment. In E. Donchin and D.B. Lindsley (Eds.) Averaged evoked potentials: Methods, results and evaluations. Washington, D.C., NASA SP-191, 1969b, 262-275.

Chapman, R.M. Evoked potentials of the brain related to thinking. In F.J. McGuigan and R. Schoonover (Eds.), Psychophysiology of Thinking. New York: Academic Press, 1973, 69-108.

Chapman, R.M. Latent components of average evoked brain responses functionally related to information processing. In International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts. Brussels: Presses Universitaires de Bruxelles, 1974a, 38-42.

Chapman, R.M. Semantic meaning of words and average evoked potentials. In International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts. Brussels: Presses Universitaires de Bruxelles, 1974b, 43-45.

Chapman, R.M. Hemispheric differences in average evoked potentials to relevant and irrelevant visual stimuli. In International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts. Brussels: Presses Universitaires de Bruxelles, 1974c, 45-46.

Chapman, R.M. (Chair.) ERPs and Language. Transcript of panel at Fourth International Congress on Event Related Slow Potentials of the Brain (EPIC IV), David A. Otto, Program Chairman, 4-10 April 1976, Hendersonville, N.C.

Chapman, R.M. Light intensity and the "storage" component in visual evoked potentials. Electroencephalog. and Clin. Neurophysiol., 1977, 43, 778.

Chapman, R.M. Connotative meaning and average evoked potentials. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior. New York: Plenum Press, in press.

Chapman, R.M. On the road to specific information in evoked potentials. In J.C. Armington, J. Krauskopf and B. Wooten (Eds.), Visual Psychophysics: Its Physiological Basis. New York: Academic Press, in press, 469-480. Chapman, R.M. and Bragdon, H.R. Evoked responses to numerical and non- numerical visual stimuli while problem solving. Nature, 1964, 203, 1155- 1157.

Chapman, R.M. and Chapman, J.A. The General Automation 18/30 as a system for the general analysis and acquisition of data in physiological psychology. Behav. Res. Methods and Instrum., 1972, 4, 77-81.

Chapman, R.M., McCrary, J.W., Bragdon, H.R. and Chapman, J.A. Latent components of event-related potentials functionally related to information processing. In J.E. Desmedt (Ed.) Progress in Clinical Neurophysiology, Vol. 6: Cognitive Components in Cerebral Event-related Potentials and Selective Attention. Basel: Karger, in press.

Chapman, R.M, McCrary, J.W., Chapman, J.A. and Bragdon, H.R. Brain responses related to semantic meaning. Brain and Language, 1978, 5, 195-205.

Chapman, R.M., McCrary, J.W. and Chapman, J.A. Short-term memory: The "storage" component of human brain responses predicts recall. Science, in press, 1978.

Dixon, W.J. (Ed.) BMDP Biomedical Computer Programs. Berkeley: Univ. of California Press, 1975.

Donchin, E., McCarthy, G., and Kutas, M. Electroencephalographic investigations of hemispheric specialization. In J.E. Desmedt (Ed.), Progress in Clinical Neurophysiology, Vol. 3: Language and Hemispheric Specialization in Man: Cerebral Event-Related Potentials. Basel: Karger, 1977, 212-242.

Halliday, A.M., Barrett, G., Halliday, E. and Michael, W.F. The topography of the pattern-evoked potential. In J.E. Desmedt (Ed.), Visual Evoked Potentials in Man: New Developments. Oxford: Oxford Univ. Press, 1977, 121-133.

Jewett, D.L., Romano, M.N. and Williston, J.S. Human auditory evoked potentials: Possible brain stem components detected on the scalp. Science, 1970, 167, 1517-1518.

ACKNOWLEDGEMENT

Supported in part by PHS grants EY01593 and EY01319 and ONR contracts N00014-77-C-0037 and CNA SUB N00014-76-C-0001. We gratefully acknowledge the aid and comments of Henry R. Bragdon, John A. Chapman, Samuel A. Shelburne, Jr., Don Montabana, and Janice K. Martin.

In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior. New York: Plenum Press, in press.
(Based on a Conference, May 1977)

CONNOTATIVE MEANING AND AVERAGED EVOKED POTENTIALS

Robert M. Chapman

Psychology Department and Center for Visual Science

University of Rochester

Rochester, N.Y. 14627

ABSTRACT

The effects of two kinds of experimental manipulation of semantic meaning were studied in Evoked Potentials (EPs), brain responses recorded from scalp monitors. Both kinds of semantic manipulation were based on Osgood's rating analyses which described three primary dimensions of connotative meaning: Evaluation, Potency, and Activity (E, P, and A). One kind of experimental variable was the semantic class of the stimulus word (E+, E-, P+, P-, A+, A-). The other kind of experimental variable was the semantic dimension of the rating scale (E, P, A) which the subject used to make semantic judgments about the stimulus words. These variables were experimentally combined in that for each trial the subject used a designated semantic scale to judge a specified stimulus word while brain activity was recorded. Using multivariate procedures, both stimulus word class and scale dimension effects on the EPs were found. Individual subject analyses demonstrated the generality of the results by showing successful discrimination of word classes and scale dimensions for each of the ten subjects analyzed separately.

Supported in part by NIH Research Grant 5 R01 EY01593 and the Advanced Research Projects Agency (Contract N00014-77-C-0037). The author thanks John W. McCrary and John A. Chapman for their collaboration in this research and Janice K. Martin for her assistance.

INTRODUCTION

Although a relatively young field, the study of language and evoked potentials is gaining momentum and sophistication (for review: Chapman, 1976; Chapman, in press, a). The work in this field is particularly difficult since linguistic problems as well as problems inherent in EP research need to be considered. A central problem involves distinguishing language effects per se from other effects, such as lower order sensory and motor effects, as well as higher order effects such as general states and cognitive processes. One strategy is to systematically relate EP effects to intra-linguistic variation within the conceptual framework provided by one of the well-delineated subfields in linguistics.

In order to investigate brain responses related to semantic meaning, we extended the technique of averaging the EEG to averaging EPs across a number of words belonging to the same semantic class (Chapman, 1974b; Chapman, Bragdon, Chapman, and McCrary, 1977). With the aid of a quantified theory of connotative semantic meaning, we found brain activity from the human scalp which is related to semantic meaning. In order to control commonly confounding variables, the subject's task was held constant, the presentation sequences were randomized, and the semantic classes were represented by a relatively large number of different words in two lists. With regard to the specificity of the linguistic effects, six different semantic classes were distinguished.

We specified and controlled internal semantic meaning using the conceptions and materials provided by Osgood's analyses of semantic meaning (Miron and Osgood, 1966; Osgood ,1971; Osgood, May, and Miron, 1975). Those analyses indicate that the connotative meaning of a word may be represented by its position in a space spanned by three semantic dimensions: Evaluation, Potency, and Activity (E, P, and A). We selected words (Heise, 1971) which are relatively "pure" in the sense that they score high or low on one of the dimensions and are relatively neutral on the other two. Thus, we used six semantic meaning classes (E+, E-, P+, P-, A+, A-) representing the positive and negative extremes of the Evaluation, Potency, and Activity dimensions.

The degree of specificity of language effects found in EPs depends on the dimensionality of the EP measures themselves (Chapman, in press, b). It is helpful to use EP measures which can focus on linguistic parts of EPs. Two possible techniques are: (1) to use the difference between EPs with and without the particular linguistic processing; and (2) to use multivariate statistical analyses which take into account all of the time points within the EPs as well as their relationships. Thus, the dimensionality of the interpretations of linguistic specificity is limited by the dimensionality of the EP measures and the dimensionality of the experimental design.

In this paper new data relating EP effects to the semantic dimension of the subjects' task (semantic differential) are given after briefly reviewing Osgood's analysis of connotative meaning and our previous results relating EP effects to connotative classes of stimulus words.

Osgood's Analysis of Connotative Meaning

The work of Osgood and his associates is an exemplar of the psychophysics of semantic meaning (e.g. Osgood, 1952; Osgood, 1971). Their work has led to the idea that connotative meaning space can be reasonably spanned by three dimensions. Thus, any connotative semantic meaning may be specified by three numbers which represent the amount of three components "in" the stimulus (usually a word).

Their analysis used semantic differential measures of meaning. The basic measure in the semantic differential technique is obtained by collecting from the subject a match between a stimulus word and a 7-point scale, defined by a pair of polar terms (e.g., good-bad). These matches were made between a large number of words and a large number of polar terms (adjective pairs). A multivariate analysis applied to these data showed a large portion of the total variance in judgments of verbal meaning could be accounted for in terms of three underlying orthogonal factors which have been called Evaluative, Potency, and Activity (E,P,A). Some of the semantic differential scales dominantly loaded on the E, P, and A semantic dimensions are schematically depicted in Figure 1.

These semantic differential techniques have been applied to 23 different language/culture groups around the world. Although the words were different and translated words may occupy different positions in the three-dimensional connotative meaning space, the analyses repeatedly derived the same E, P, and A dimensions for spanning those meaning spaces. These cross-cultural analyses as a whole suggest that human beings share a common framework within which they allocate concepts in terms of their semantic meanings. This communality overrides gross differences in both language and culture.

This quantitative system does not deal with denotative meaning per se which would appear to have many more dimensions. Rather it deals with connotative (affective) aspects of semantic meaning. In color measurement, trichromatic specification says little about

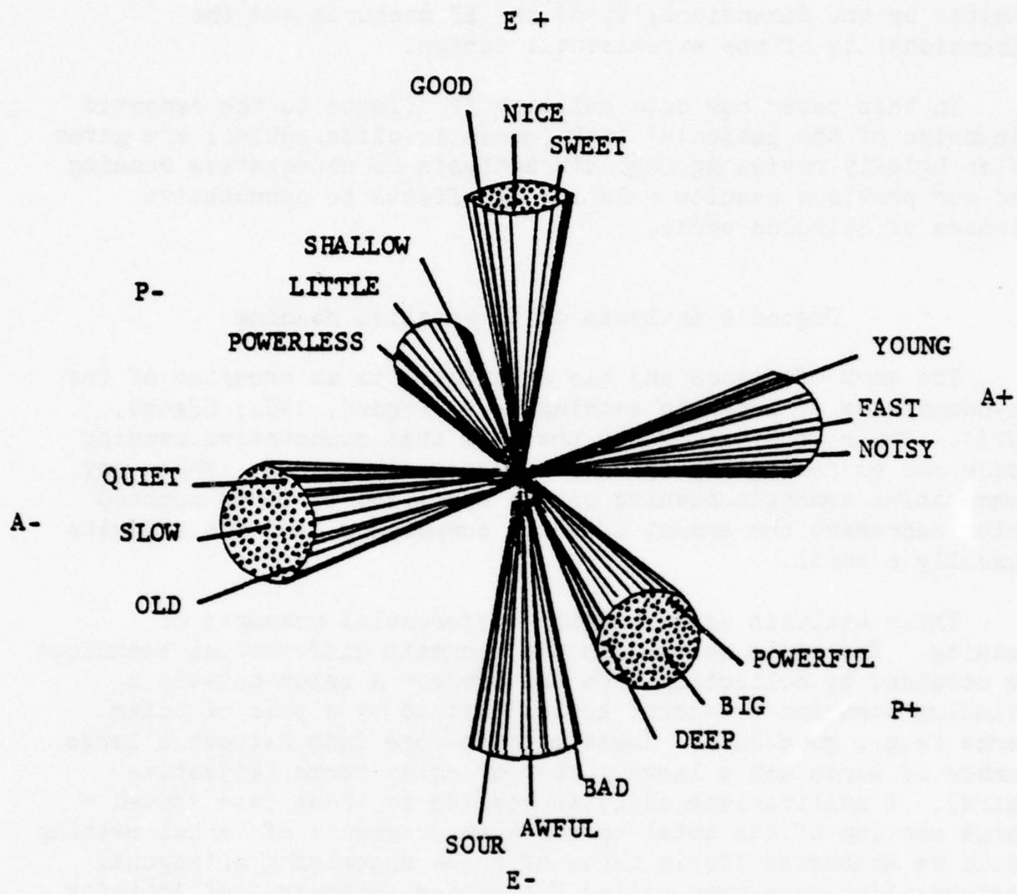


Figure 1. The Evaluation (E), Potency (P), and Activity (A) structure of connotative meaning. Some of the semantic differential scales dominantly loaded on E, P, and A. Based on Osgood (1964).

spatial patterns although the same visual stimuli involve both spatial and color aspects. Similarly, tri-connotative specification of semantic meaning says little about denotative meaning.

There is available, then, Osgood's well-defined, objectively-measured, widely-tested, fundamental analysis of semantic meaning. It enables one to work within a domain which is explicitly delimited, which has dimensions that are quantified, and which readily lends itself to objective replication.

Evoked Potentials and Connotative Meaning

When a stimulus word is presented, it evokes a number of neural processes, some of which are concerned with meaning. The detection of semantic meaning in EPs permits a more direct examination of language and its neurophysiological processes, and this opens new areas of research and application. Our data are encouraging in that significant effects related to Osgood's semantic dimensions have been shown (Chapman, 1974b; Chapman, Bragdon, Chapman, and McCrary, 1977; Chapman, McCrary, Chapman, and Bragdon, in press). Other work has also indicated effects (Begleiter, Gross and Kissin, 1967; Begleiter, Gross, Projesz and Kissin, 1969; Begleiter and Platz, 1969).

For our research we selected words on the basis of Osgood's Evaluative, Potency and Activity dimensions of connotative meaning (Heise, 1971). Six semantic meaning classes (E+, E-, P+, P-, A+, A-) representing the positive and negative extremes of each of the three dimensions were used. Twenty words from each of the six semantic classes were randomly assigned to a list. Two such lists were constructed with different words, except for the P- category where the same words were used. The words belonging to these semantic meaning classes were visually presented and the average EPs for these classes were analyzed. The physical parameters of the stimuli (various spatial characteristics) vary from one word to the next but the physical parameters tend toward the same average for the various groups of words (Chapman, McCrary, Chapman, and Bragdon, in press). Using two lists provided an additional control. While the background EEG is averaged to obtain EPs, the physical characteristics of the words are averaged to control for their effects and the meanings of the words are averaged to provide a common core of connotative meaning. The words within each list were given in different random orders from run to run, so that the subjects could not anticipate either a semantic class or a particular word. Thus, differences in the EPs to these semantic categories can be associated with post-stimulus processing of semantic information, with the comparison of responses to the two lists helping establish the reliability and generality of the effects. Because the brain responses to be compared were derived from semantic categories which are randomly interspersed, it is difficult to attribute the obtained differences to anything other than semantic processing or effects arising from semantic processing.

In our initial research on semantic meaning (Chapman, 1974b; Chapman, Bragdon, Chapman, and McCrary, 1977) we used a scoring template approach to compare EPs to word classes from opposite ends of Osgood's dimensions. For the scoring template for the Evaluative dimension, the average EP (from CPZ) for E- words was subtracted from that of E+ words, averaged over three subjects on two word lists. This scoring template was then used to measure each EP by computing the Pearson product-moment correlation coefficient using the 102 corresponding time points of the scoring template and the EP. This yielded a single measure for each EP reflecting its similarity to the scoring template. Using this measure significant differences were found between EPs for E+ and E- word classes. (The EP template measures were z-transformed [arc-tanh] before applying t-tests for correlated measures). The t values for all 12 subjects were in the predicted direction, i.e., positive. For the three subjects involved in the development of the scoring template, 81% of their EPs were correctly classified into E+ or E- word classes on the basis of the relative magnitudes of their correlations with the E template. A somewhat smaller, but significant, success rate was obtained for the nine subjects in the independent cross-validation group. P (Potency) and A (Activity) templates derived in the same way had somewhat lower success rates in discriminating P+ from P- and A+ from A- word classes, respectively. The relative strengths of the EP effects found for the E, P, and A dimensions might be expected from Osgood's analysis. Evaluation has been found to be the most pervasive aspect of connotative meaning, followed by the Potency and then the Activity dimensions. The use of a scoring template to measure EPs for semantic effects was an exploratory technique.

Encouraged by these template results we have continued our research on semantic meaning and EPs with the aid of multivariate statistical techniques (Chapman, 1976; Chapman, McCrary, Chapman and Bragdon, in press; Chapman, in press, a, b). One of the problems was coping with the large individual differences in EP waveforms. These overall waveform differences, while not the semantic effects of interest, were relatively stable characteristics of each individual subject. This problem was solved by standardizing the EPs for each subject separately (transforming to z-scores at each time point) before proceeding with the analysis. A varimaxed principal components analysis was computed on the standardized EPs from a group of 10 subjects in order to obtain component scores. These EP component scores were used in a multiple discriminant analysis to develop classification functions for the six semantic word classes. The success rates in classifying EPs to the semantic classes were significantly better than chance. Classification functions were developed separately for the EP data from each list of words and the results cross-validated by several procedures: (i) jackknifed (one-left-out procedure), (ii) other word list, and (iii) new subject.

When the EPs were classified to word classes from opposite ends of each semantic dimension separately (E+ vs. E-, P+ vs. P-, A+ vs. A-), the average apparent success rate was 97% and the jackknifed cross-validation success rate was 90% (chance was 50%).

When the same classification functions were applied to the EP data obtained from the other word list, the overall sucess rate was 73%.

Multidimensional analyses considered the EP data for all three semantic dimensions at once, in which case six semantic classes were discriminated from each other (E+, E-, P+, P-, A+, A-). The classification rates were significantly better than chance. Overall, the jackknifed success rates (where each EP is left out of the development set and then classified) were 42% for List 1 and 43% for List 2 data, some 2.5 time better than chance (16.7%). The other-list cross-validations averaged 40%. Thus combinations of components of these EPs were powerful detectors of semantic differences.

It is to be noted that all of the above success rates were obtained across subjects. That is, the same classification functions were used for all ten subjects. This is evidence that not only can EP effects be found that relate to connotative semantic meaning but these EP effects tend to be the same in different individuals.

A further test of the generalizability of the findings was made by applying the classification functions to a new subject, one not used in developing the analysis. After standardizing his EPs and using component scoring and discriminant functions developed from the separate group of 10 subjects, 42% of the new subject's EPs were correctly classified into the six semantic classes, essentially the same rate as the jackknifed accuracy of the group of 10 subjects and significantly better than chance (16.7%).

SEMANTIC-DIFFERENTIAL SCALES AND SEMANTIC WORD CLASSES

In the results summarized above, the subject's task was simply to repeat each word aloud after it was flashed. It was of interest, for several reasons, to change to a semantic-differential judgment task, one in which the subject makes a judgment about each word on a designated bipolar adjective scale. This was the task that Osgood used to develop his semantic data and quantitative information about the loadings of various scales on Osgood's dimensions is available. This made it possible to select judgment scales that strongly represented each of the E, P, and A semantic dimensions.

In our previous research internalized representation of semantic meaning was manipulated by carefully selecting stimulus words. Another aspect of internalized representation may relate to an individual's semantic expectancies. When the same word is presented on different occasions, a subject may be seeking different kinds of semantic information. That is, a subject may have various kinds of semantic expectancies and, consequently, the semantic information in the words may be processed along various semantic dimensions. For example, an individual might be primarily concerned with potency (powerful-powerless) when a stimulus word "official" occurs or he might be primarily concerned with evaluation (good-bad). Do the Evoked Potentials related to the word "official" vary for these different semantic expectancies? Do these different semantic expectancies have their own EP effects?

In order to study questions of this sort, we manipulated the semantic expectancy by assigning various semantic differential scales to the subjects at different times (Table I). The subject's task was the semantic differential task, as used by Osgood in developing his semantic analysis.

A further reason to change the subject's task from repeating the word to giving a numerical judgment (+3 to -3) was as an additional control for speech effects. The same vocalizations were made to all word classes, as well as for all scale dimensions.

Thus, this research studied two kinds of experimental manipulation of semantic meaning: word class of the stimulus word (E+, E-, P+, P-, A+, A-), and scale dimension (E, P, A) which the subject used to make semantic differential judgments about the stimulus words (Fig. 2). Five bipolar scales that were heavily loaded on (correlated with) each of Osgood's semantic dimensions (E, P, and A; see Table I) were selected (Osgood, 1964). Each of these 15 scales was used with each stimulus word. Thus, the effects of two kinds of experimental manipulation of semantic meaning were studied: (1) the semantic class of the stimulus word, and (2) the dimension of the semantic scale (E, P, A) which the subject used to make semantic-differential judgments about the stimulus words. These variables were experimentally combined in that for each trial the subject used a designated semantic scale to judge a specified stimulus word. Separate analyses identified word class and scale dimension effects in the EPs at better than chance levels.

Synopsis of Procedure

During each experimental run, 120 words were flashed in random order while the subject's EEG was recorded. For each run, there were 20 words representing each of six classes of semantic meaning lying at the positive and negative extremes of each of the Osgood dimensions: Evaluation, Potency, and Activity. The subject was

Table I

Loadings of Semantic Differential Scales on Evaluation (E), Potency (P), and Activity (A) *

SCALE		E	P	A
E Dominantly				
E1	nice-awful	.96	-.02	-.09
E2	sweet-sour	.94	.02	-.04
E3	good-bad	.93	.03	-.05
E4	heavenly-unheavenly	.93	.00	-.21
E5	mild-harsh	.92	-.20	-.06
P Dominantly				
P1	big-little	-.05	.81	-.24
P2	powerful-powerless	.16	.75	.18
P3	deep-shallow	-.11	.69	-.32
P4	strong-weak	.04	.68	.13
P5	long-short	.02	.64	-.23
A Dominantly				
A1	fast-slow	-.14	.22	.64
A2	young-old	.39	-.42	.56
A3	noisy-quiet	-.39	.25	.56
A4	alive-dead	.52	.13	.55
A5	known-unknown	.16	.10	.48

----

* American English semantic differential loadings reported in Osgood, 1964. Loadings shown are for the first listed adjective of each pair. "Good", "Powerful", and "Fast" are represented by the positive poles of E, P, and A.

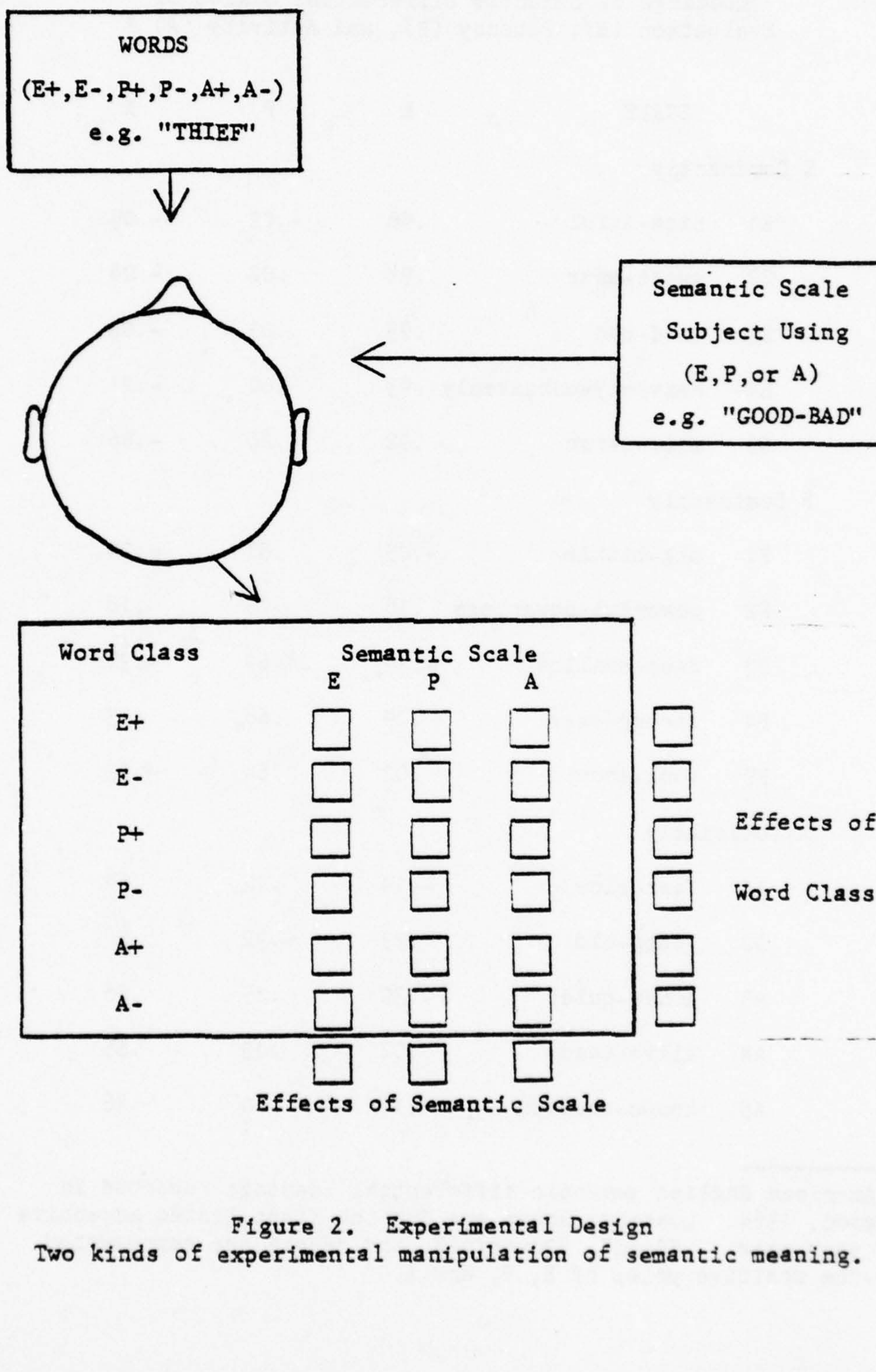


Figure 2. Experimental Design
Two kinds of experimental manipulation of semantic meaning.

assigned a particular semantic scale for use during the run in judging each word as it was presented. The EEGs for the 20 stimulus words representing each semantic class were averaged for the run to obtain the evoked potentials (EPs) used in subsequent analyses. A total of 30 such runs was required to complete the collection of 180 such averaged EPs for each individual across all experimental conditions:

(1) Six semantic classes of stimulus words,
(2) Two different lists of words (to control for specific stimulus characteristics or properties other than connotative meaning),
(3) Three semantic task dimensions, each represented by five different scales (to control for specific scale properties other than dominant semantic dimension).

METHOD

The research steps are summarized in the Flow Chart of Experiment (Table II).

The six semantic categories were represented by the same word lists used previously (Chapman, 1974b; Chapman, Bragdon, Chapman, and McCrary, 1977; Chapman, McCrary, Chapman, and Bragdon, in press). The words within each list were given in different random orders from run to run, so that the subjects could not anticipate the semantic class of the stimulus words during the experiment.

Five scales that are heavily loaded on each of Osgood's three semantic dimensions (Evaluation, Potency, and Activity) were selected (Osgood, 1964). Each of these 15 semantic scales (Table I) was used with each stimulus word. This required 15 runs with List 1 and 15 runs with List 2, making a total of 30 runs for each subject. The scales were given in different random orders for each subject.

Before each run the subject was assigned a semantic scale, e.g. "nice-awful," which he was to use on all 120 words in that run. The subject was asked to rate each stimulus word on the designated semantic scale using values from +3 to -3. The instructions to the subject when the scale was "nice-awful" were: If the meaning of the word to you is more nice than awful, then give a + rating, with a 1, 2, or 3 to express various degrees of niceness. On the other hand, if the meaning of the word to you is more awful than nice, give a - rating using 1, 2, or 3 to indicate the degree of awfulness. If the word is perfectly neutral on that scale, give a "zero." For each scale, regardless of whether it was "nice-awful," "big-little," "fast-slow," or some other scale,

Table II

FLOW CHART OF EXPERIMENT

2 LISTS OF WORDS SELECTED FOR 6 SEMANTIC CLASSES: E+, E-, P+, P-, A+, A- BASED ON OSGOOD'S 3-DIMENSIONAL ANALYSIS

5 SEMANTIC DIFFERENTIAL SCALES SELECTED FOR EACH OF 3 DIMENSIONS: E, P, A BASED ON OSGOOD'S ANALYSES

WORDS FLASHED ON CRT
EEG RECORDED
SUBJECT GIVES SEMANTIC DIFFERENTIAL

EVOKED POTENTIALS (N=20) COLLECTED FOR EACH SEMANTIC WORD CLASS WITH EACH SEMANTIC SCALE

VARIMAXED PRINCIPAL COMPONENTS ANALYSIS ON EPs OF 102 TIME POINTS, COMPONENT SCORES COMPUTED FOR EACH EP.

DISCRIMINANT ANALYSES USING COMPONENT SCORES TO CLASSIFY EPs INTO:

SEMANTIC WORD CLASSES (6)

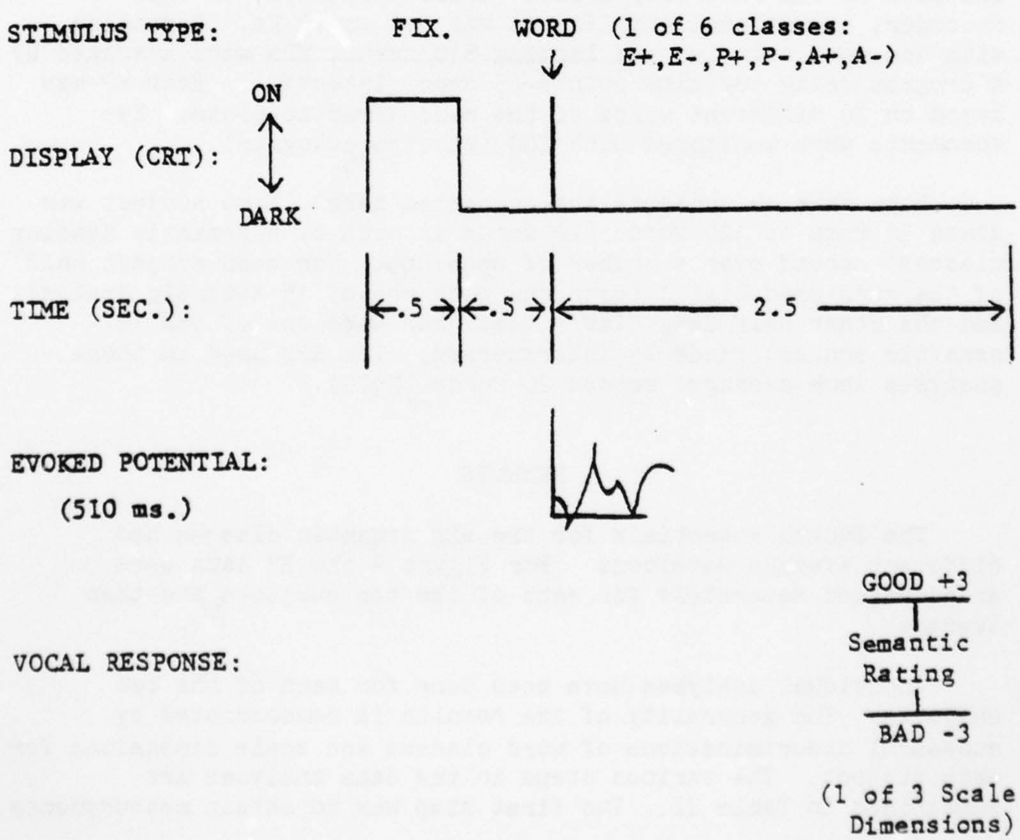
SEMANTIC SCALE DIMENSIONS (3)

numerical values from +3 to -3 were used. After each word was flashed the subject gave his semantic differential rating aloud.

A computer-generated display system presented each word as a briefly flashed stimulus on a CRT (Fig. 3). The subject sat in a dark, sound-damped chamber. The average word subtended a visual angle of 1.5 degrees with a duration of 17 msec. Each letter was formed by lighting appropriate positions in a 5 by 7 matrix. A fixation target was presented (0.5 sec. duration) one second before each word. After each word was flashed the subject gave his semantic differential rating (+3 to -3) toward the end of the 2.5 sec. interval between each word and the fixation stimulus for

Figure 3

Diagram of Single Trial.



the next trial. This task assured that each stimulus word was perceived and provided access to a behavioral measure. The brain activity following these word stimuli was averaged separately for each of the semantic meaning classes in conjunction with each semantic scale. The sequence for each word presentation (a trial) within each run was as follows:

(1) Fixation target on for 0.5 sec.
(2) Blackout for 0.5 sec.
(3) Stimulus word flashed (approximately 17 msec.)
(4) Blackout for 2.5 sec., during which time the subject gave a number representing his semantic judgment of the word on a designated scale.

An experimental run consisted of 120 words presented in this fashion.

During experimental runs, the subject's EEG was picked up from standard Grass electrodes (silver cup shape) which were attached by bentonite CaCl paste. The data reported here were recorded from a scalp location one-third of the distance from CZ to PZ (CPZ recorded monopolar to linked earlobes). The frequency bandpass of the recording system (Grass polygraph, FM tape recorder, operational amplifiers) was 0.1 to 70 Hz. Beginning with the word stimulus and lasting 510 msec., EPs were averaged by a program using 102 time points (5 msec. interval). Each EP was based on 20 different words of the same semantic class. Eye movements were monitored with EOG (electrooculogram).

Data from 10 subjects are presented here. Each subject was given 30 runs of 120 words (20 words in each of 6 semantic meaning classes) spread over a number of sessions. For each subject half of the runs used List 1 (each run with one of 15 semantic scales) and the other half used List 2 (each run with one of the 15 semantic scales) randomly interspersed. The EPs used in these analyses were averages across 20 words (N=20).

RESULTS

The Evoked Potentials for the six semantic classes had different average waveforms. For Figure 4 the EP data were standardized separately for each of the ten subjects and then averaged.

Individual analyses have been done for each of the ten subjects. The generality of the results is demonstrated by sucessful discriminations of word classes and scale dimensions for each subject. The various steps in the data analyses are summarized in Table II. The first step was to obtain measurements

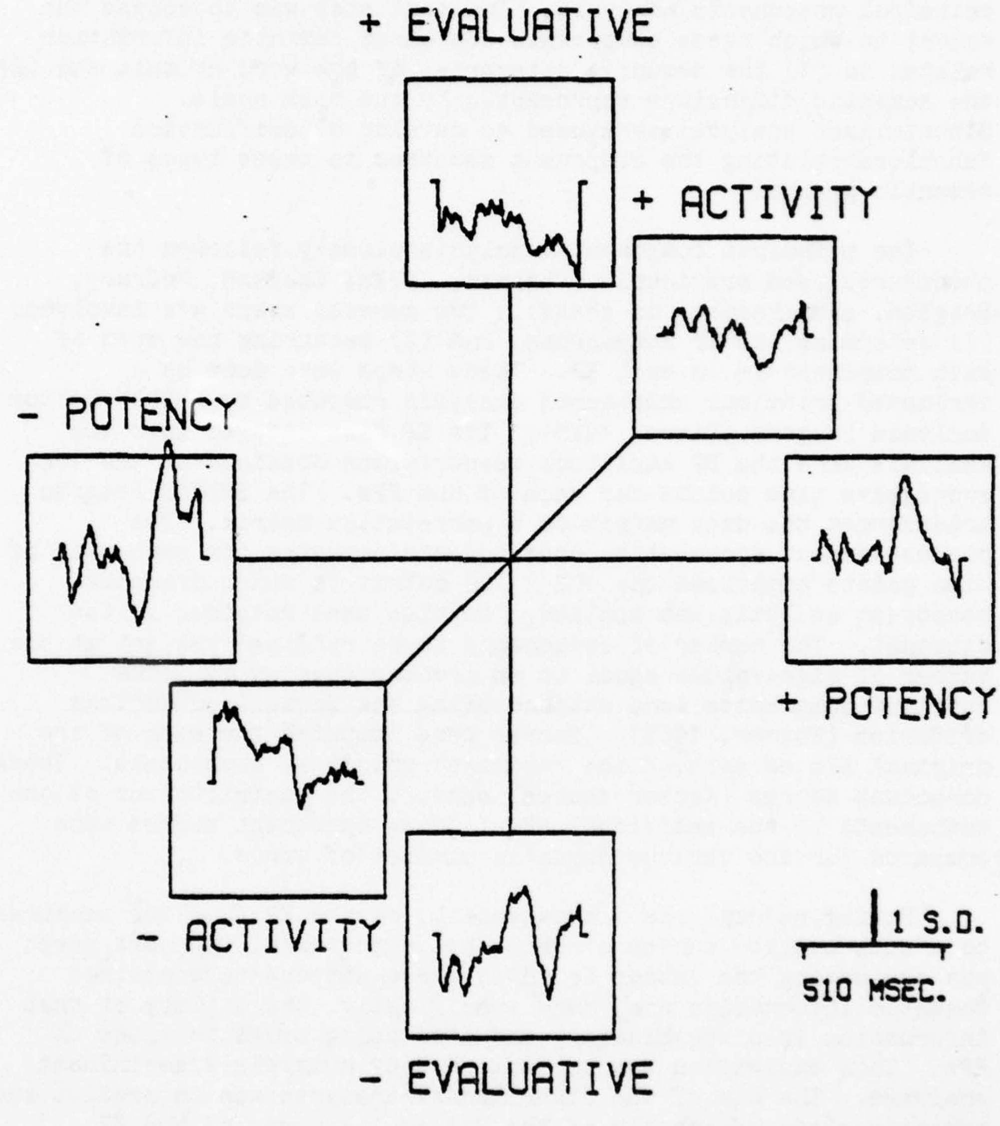


Figure 4. Average Evoked Potentials (EPs) for six semantic classes after standardization. The semantic word classes are based on Osgood's Evaluation, Potency, and Activity dimensions which define a three-dimensional connotative meaning space, represented schematically here. The EPs cover 510 msec (102 time points X 5 msec) along the horizontal, beginning at the time the words were flashed. The vertical axes for the EPs are in standard units (z scores). For the Standardized Potentials each subject's data at each time point were transformed to z scores (means=0 and standard deviation=1). Averages include data for two word lists and ten subjects. Monopolar recordings (bandpass: 0.1 to 70 Hz) from a scalp location 1/3 of the distance from Cz to Pz. Positive is up.

of components of the brain potentials by use of a varimaxed principal components analysis. The next step was to assess the extent to which these components contained semantic information related to (1) the semantic categories of the word stimuli and (2) the semantic dimensions represented by the task scale. Discriminant analyses were used to develop classification functions relating the component measures to these types of semantic groups.

The principal components analysis closely followed the procedures used previously (Chapman, 1974a; Chapman, McCrary, Bragdon, and Chapman, in press). Two general steps are involved: (1) determing the EP components, and (2) measuring how much of each component is in each EP. These steps were done by a varimaxed principal components analysis computed by BMDP4M Factor Analysis Program (Dixon, 1975). The EP data entered into the analysis were the EP amplitude measurements obtained at the 102 successive time points for each of the EPs. The BMDP4M Program transformed the data matrix to a correlation matrix. The product-moment correlation coefficients computed for each pair of time points comprised the 102 x 102 matrix to which principal component analysis was applied. Unities were retained in the diagonal. The number of components to be retained was set at the number of eigenvalues equal to or greater than unity. The retained components were rotated using the normalized varimax criterion (Kaiser, 1958). Scores were computed for each of the original EPs on each of the varimaxed principal components. These component scores (factor scores) measure the contributions of the components to the individual EPs. These component scores were compared for the various semantic classes of words.

Having reduced the dimensionality of the EP from 102 measures to a much smaller number of principal components, the next step was evaluating the extent to which these components contained semantic information and, more specifically, the utility of that information in discriminating and predicting semantic class of EPs. This evaluation was accomplished by multiple discriminant analyses. The aim of the discriminant analyses was to predict the semantic class membership of the EPs on the basis of the EP measures (component scores). The discriminant analyses were done by the BMDP7M Stepwise Discriminant Analysis Program (Dixon, 1975). This program was applied to the component scores derived from the principal components analyses. A set of linear classification functions was computed by choosing the independent variables in a stepwise manner. Using these functions, the probabilities of each EP belonging to each semantic class were computed.

Two separate multiple discriminant analyses, one for each word list, were performed on each subject's EP data to determine the ability of the EP component measures to discriminate simultaneously among all six of the semantic classes of stimulus words. The success rates of classifying EPs into the appropriate semantic classes were averaged for the two word lists and are presented in Table III.

The overall success rate (pooling lists and subjects) in classifying EPs involved in the computation of the discriminant analyses and classification functions was 43.5 percent. The success was well beyond the chance level of 16.7 percent. These results were cross-validated by two procedures: (1) jackknifed cross-validation and (2) other-list cross-validation. The jackknifed procedure assesses the classification success when EPs are left out of the development set one at a time and the discriminant functions so developed are used to classify the EPs as they are left out. This technique is used to estimate the success which would be expected in classifying other, additional EPs obtained using the development list. An overall success rate of 31.0 percent was obtained with this procedure. In the other-list cross-validation, the classification rules developed for EPs obtained with one word list are used to classify EPs collected with the other list of word stimuli. This provides a further check on generalizability of the discriminant functions and tests their likely success rate in classifying other, additional EPs obtained using a different set of words. As shown in Table III, the overall accuracy in classifying such other-list EPs was 26.8 percent for these ten subjects.

Since all six semantic classes of stimuli were represented simultaneously in these analyses, the success rate expected by chance was 16.7 percent. The success rates were all well beyond this chance level (chi-squares in Table III).

In addition to semantic class of the stimulus words, the semantic dimension of the subjects' task was investigated. The average EP data for E, P, and A semantic differential tasks are shown in Figure 5 as Standardized Potentials. An additional discriminant analysis was performed for each of the ten subjects to evaluate the extent to which the EP component measures also contain information about the semantic nature of the subject's task (Table III, 3 Scale Dimensions). The specific aim of the analysis was to determine whether functions of these EP components could be developed to differentiate among EPs according to the semantic dimension of the scale being used by the subject to make judgments about the stimuli being presented. The overall success rate of these functions in correctly classifying the EPs used in their development was 47.4 percent. This rate of success was better than the chance rate of 33.3 percent. The jackknifed cross-validation, using the one-left-out procedure described

Table III

Percentages of EPs Correctly Classified

6 Semantic Groups of Words
3 Semantic Dimensions of Scales
2 Word Lists

	6 Semantic Groups Multi-Dimensional Analysis			3 Scale Dimensions	
Subject	Development	Cross-Validation Jackknifed	Cross-Validation Other List	Development	Jackknifed Cross-Validation
A	46.1	31.6	30.6	54.4	50.6
B	36.1	28.9	23.3	47.8	41.7
C	57.2	38.4	30.0	47.2	47.2
D	38.4	28.4	23.9	45.0	38.9
E	35.6	24.4	28.9	43.3	40.6
F	39.4	31.7	13.8	47.2	44.4
G	40.6	32.2	27.2	46.1	42.2
H	43.3	30.6	25.6	45.0	42.8
I	48.9	30.0	31.7	48.9	46.1
J	49.4	33.4	32.8	48.9	44.4
OVERALL	43.5	31.0	26.8	47.4	43.9
CHANCE EXPECTATION	16.7	16.7	16.7	33.3	33.3
CHI-SQUARE df = 1	931.2	263.2	131.8	159.4	89.8

Each individual percentage based on 180 EPs.

All values of Chi-square corrected for discontinuity.

Chi-square (df=1, p=.001) = 10.8

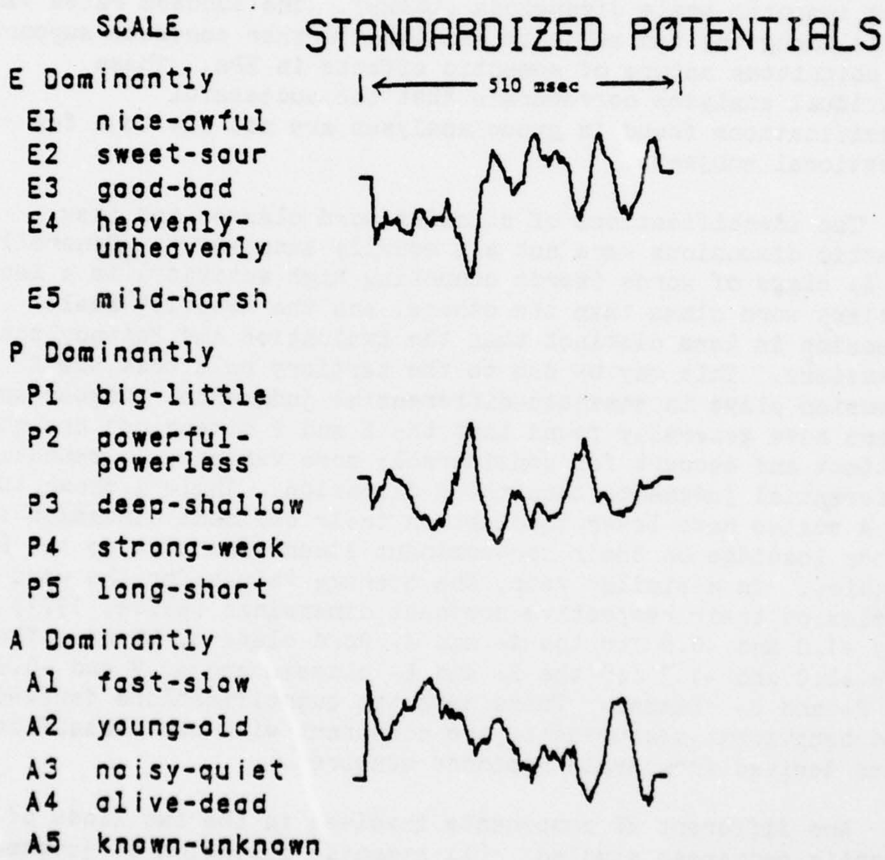


Figure 5. Standardized Evoked Potentials (from CPZ) for semantic differential task scales which are dominantly loaded on Evaluation (E), Potency (P), and Activity (A) semantic dimensions. Data averaged across stimulus word classes, two word lists, and ten subjects. See Figure 4 legend for information about Standardized Potentials. The vertical scale is indicated by the peak-to-peak amplitude of 0.26 z-score units for the response to E scales.

previously, resulted in the correct identification of the semantic dimension of the task scale of 43.9 percent of the EPs. This is an indication of the likely success to be obtained in classifying other, additional EPs obtained with the subjects while using these semantic differential scales. The chi-square statistics indicate that these rates of correct classifications are well beyond chance expectations.

The individual analyses (Table III) indicate that EP data from each of the subjects could be used individually to discriminate successfully among semantic word groups (stimuli) and among semantic scale dimensions (tasks). The success rates varied little among the ten subjects and lend further concrete support to the ubiquitous nature of semantic effects in EPs. These individual analyses corroborate that the successful classifications found in group analyses are not due to a few exceptional subjects.

The identifications of stimulus word classes and task semantic dimensions were not all equally successful. Generally, the A+ class of words (words connoting high activity) is a less distinct word class than the others, and the Activity scale dimension is less distinct than the Evaluation and Potency scale dimensions. This may be due to the tertiary role that the A dimension plays in semantic-differential judgments. Osgood and others have generally found that the E and P dimensions are more distinct and account for considerably more variance in semantic differential judgments than the A dimension. Table I shows that the A scales have lower loadings on their dominant dimension and higher loadings on their non-dominant dimensions than do the E or P scales. In a similar vein, the average values for the word classes on their respective dominant dimensions (Heise, 1971) were only +1.0 and -0.8 for the A+ and A- word classes, whereas they were +2.0 and -1.3 for the E+ and E- classes and +1.9 and -0.6 for the P+ and P- classes. These semantic quantifications derived from behavioral measurements are consonant with our classification rates derived from brain response measures.

Are different EP components involved in the two kinds of semantic processes studied: (i) semantic dimension of judgement scale and (ii) connotative meaning of stimulus words? Or are these similar phenomena in terms of their EP effects? Three discriminant analyses were available for each subject: one discriminating among the three task scale dimensions and two (one for each word list) differentiating among the six semantic classes of word stimuli. The first EP component to enter each of these discriminations was noted for each subject and frequency counts were made of how often the EP component entered (1) was the same for the two stimulus word class discriminations and (2) was the same for the task scale dimension discrimination and either or both of the word class discriminations (Fig. 6). In 40% of the pairs of stimulus word discriminations, the first EP components were identical. However, the first component entering the discrimination of task scale dimensions matched those entering either of the stimulus word discriminations only once out of 20 possible matches. The difference is statistically reliable (Fisher's exact probability=.03). The first two EP components entering each discriminant analysis were also compared. They were identical on 50% of the possible occasions for the two word lists (discriminating semantic word classes). The first two scale components matched those in either of the word list analyses 15% of the possible times (6 out of 40). These differences in

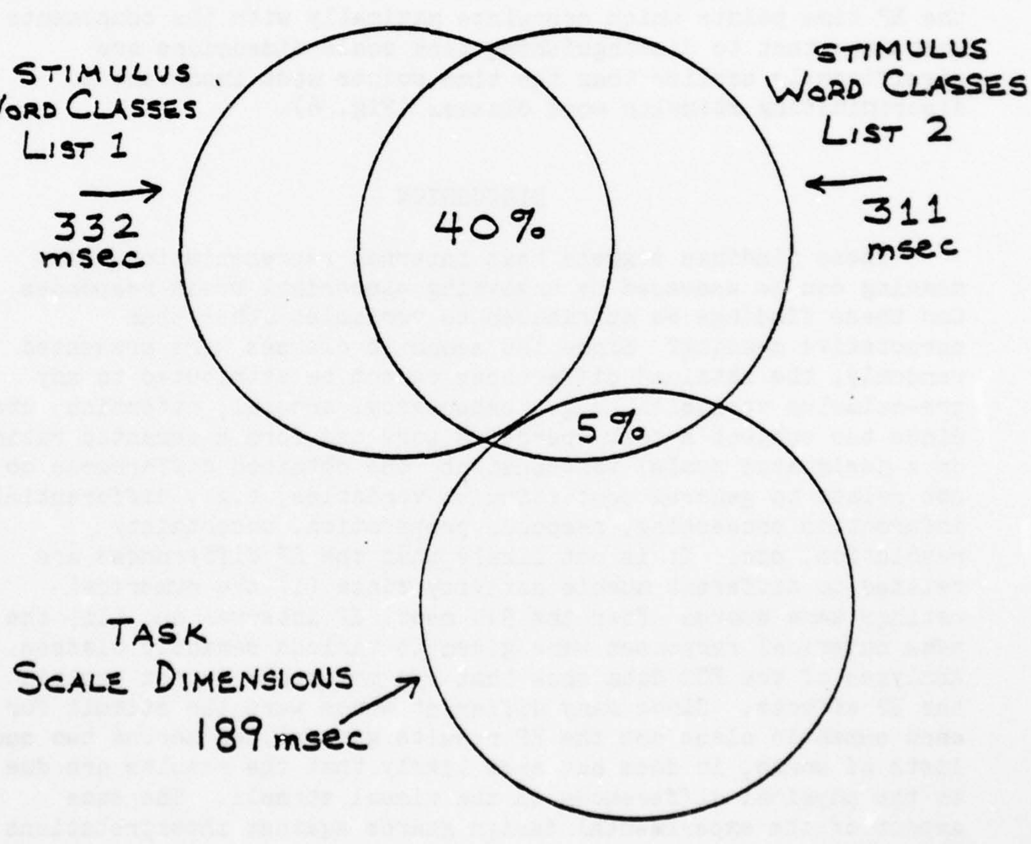


Figure 6. Overlap among first EP components to enter each of three kinds of discriminant analyses and mean latencies of maximum loadings of first EP components to enter.

frequencies of matches are statistically reliable (corrected chi-square=7.43, df=1, p<.01). Thus, EP components contributing most to distinguishing among stimulus word classes are seldom those which contribute most to distinguishing task scale dimensions.

In a different approach to the same question, the EP latencies which correlated maximally (maximum loading) with the first EP component entering each of the discriminations were tabulated. The mean latencies for the separate word list analyses, 332 msec. and 311 msec., did not differ significantly (t=.62, df=9). The mean of such latencies for the first components entered in discriminating task scale dimensions was 189 msec., which differed reliably from the mean latency for the stimulus word class discriminations (t=3.15, df=9, p<.02). Thus, the EP time points which correlate maximally with the components most important to distinguishing task scale dimensions are significantly earlier than the time points most important to discriminating stimulus word classes (Fig. 6).

DISCUSSION

These findings suggest that internal representations of meaning can be assessed by analyzing electrical brain responses. Can these findings be attributed to variables other than connotative meaning? Since the semantic classes were presented randomly, the obtained differences cannot be attributed to any pre-stimulus variables, e.g., expectancy, arousal, attention, etc. Since the subject's task (perceive word and form a semantic rating on a designated scale) was constant, the obtained differences do not relate to general post-stimulus variables, e.g., differential information processing, response preparation, uncertainty resolution, etc. It is not likely that the EP differences are related to different muscle activity since (i) the numerical ratings were spoken after the 510 msec. EP interval and (ii) the same numerical responses were given to various semantic classes. Analyses of the EOG data show that eye movements do not explain the EP effects. Since many different words were the stimuli for each semantic class and the EP results generalized across two such lists of words, it does not seem likely that the results are due to the physical differences in the visual stimuli. The same aspect of the experimental design guards against interpretations based on surface linguistic features. Finally, distinguishing six semantic classes indicates a degree of specificity which generally taxes interpretations in terms of variables other than connotative meaning.

Previous research investigated EP effects associated with the same six semantic classes of words when the subjects's task was merely to repeat each word after it was flashed (Chapman, 1974; 1976; in press, a; Chapman, Bragdon, Chapman and McCrary, 1977; Chapman, McCrary, Chapman and Bragdon, in press). In the present experiment the subject's task was to give semantic differential ratings of each word on semantic scales predominantly loaded on one of three semantic dimensions. Does the increased task complexity prevent discriminating the word class by brain response measures? Does the use of different scales, loaded on different semantic dimensions, interfere with identifying the word class of stimulus words? Do the various semantic expectancies engendered by prior assignment of semantic scales interfere with identifying the stimulus word classes?

The present results indicate that semantic effects of stimulus words continue to be detectible in EPs when the subject is engaged in a semantic task considerably more complex than only repeating the stimulus words. The added complexity of the experimental conditions clearly does not obscure the semantic word effects.

In addition, the results provide evidence that EP effects may also be used to discriminate among semantic expectancies, sets or contexts (E, P and A scale tasks) regardless of the semantic location of the stimulus words (E+, E-, P+, P-, A+, A-). Semantic judgements were elicited from the subjects using 15 scales selected to represent the E, P, and A dimensions. The subjects' internal semantic events were manipulated by the subjects' task which is set prior to delivering the stimulus word. In this sense, the task provides a semantic context or expectancy within which the stimulus word is to be evaluated. We are not using semantic expectancy here to mean the subjects' expectancy of a particular stimulus word or word class (which were randomized), but rather to mean the subjects' previously established context (delineated by dominant dimension of semantic scale) which the subject expects to apply to flashed stimulus words. The task scale dimension variable was manipulated in this experimental design independently of the stimulus word class. It was not previously known whether the task scale dimensions would have distinctive effects in EPs and, if so, whether these effects would interact with those associated with stimulus word class. The present results indicate that the semantic context established by various scales does have its own EP effects, which do not appear to interact with detection of stimulus word class.

In a manner which parallels our conclusions about identifying stimulus word class, there is some generality to identifying task scale dimension. A number of semantic differential scales were used to represent each semantic dimension (five adjective pairs for each) in order to establish general relationships to EPs, not tied to particular exemplars of the semantic scales. This parallels the use of many exemplars of stimulus word class in establishing the generality of those EP effects.

The scale dimensions could be identified by separate analyses of each individual's data (Table III). The success of these analyses supports the universality of the EP effects across individuals.

Analyses of the EP components involved in the discriminations indicate that the components reflecting the greatest differences among task scale dimensions are different from the EP components which discriminate maximally among semantic classes of stimulus words. The maximal representation of effects in the EPs occurs significantly earlier for the task scales than for the stimulus words. These findings support the conclusion that these are different kinds of semantic effects. Moreover, the earlier maximal representation in the EPs of the task scales fits the interpretation of a semantic expectancy established by the semantic differential scale assigned to the subject before the stimulus words are flashed. These data lead to the hypothesis that, following the presentation of each stimulus word, a process relating to the semantic differential scale used to judge the word occurs before the connotative meaning of the stimulus word is fully developed.

In general, the research provides evidence that two kinds of semantic variables can be independently and simultaneously identified in EPs: (1) the semantic class of stimulus words and (2) the semantic dimension of semantic-differential scales being used to judge stimulus words. These findings have important implications for applications as well as a basic understanding of the processes. In this experiment, two kinds of semantic effects were registered in the EP and could be used to assess different semantic aspects: (1) the processing of the semantic meaning in stimulus words, regardless of the semantic expectancies of the subject, and (2) the semantic expectancies of the subject, regardless of the semantic content of stimulus words.

REFERENCES

Begleiter, H., Gross, M.M. and Kissin, B. Evoked cortical responses to affective visual stimuli. Psychophysiology, 1967, 3, 336-344.

Begleiter, H., Gross, M.M., Porjesz, B. and Kissin, B. The effects of awareness on cortical evoked potentials to conditional affective stimuli. Psychophysiology, 1969, 5, 517-529.

Begleiter, H. and Platz, H. Cortical evoked potentials to semantic stimuli. Psychophysiology, 1969, 6, 91-100.

Callaway, E. Brain Electrical Potentials and Individual Psychological Differences. New York: Grune and Stratton, 1975.

Chapman, R.M. Latent components of average evoked brain responses functionally related to information processing. In International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts. Brussels: Presses Universitaires de Bruxelles, 1974a, 38-42.

Chapman, R.M. Semantic meaning of words and average evoked potentials. In International Symposium on Cerebral Evoked Potentials in Man, pre-circulated abstracts. Brussels: Presses Universitaires de Bruxelles, 1974b, 43-45.

Chapman, R.M. (Chair.) ERPs and Language. Transcript of panel at Fourth International Congress on Event Related Slow Potentials of the Brain (EPIC IV), David A. Otto, Program Chairman, 4-10 April, 1976, Hendersonville, N.C.

Chapman, R.M. Language and evoked potentials. In D.A. Otto (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. U.S. Government Printing Office, Washington, D.C., in press (a).

Chapman, R.M. Method of EP analysis in linguistic research. In D.A. Otto (Ed.), Multidisciplinary Perspectives in Event-Related Brain Potential Research. U.S. Government Printing Office, Washington, D.C., in press (b).

Chapman, R.M., Bragdon, H.R., Chapman, J.A. and McCrary, J.W. Semantic meaning of words and average evoked potentials. In J.E. Desmedt (Ed.), Progress in Clinical Neurophysiology. Vol. 3: Language and Hemispheric Specialization in Man: Cerebral Event-Related Potentials. Basel: Karger, 1977, 36-47.

Chapman, R.M., McCrary, J.W., Bragdon, H.R., & Chapman, J.A. Latent components of event-related potentials functionally related to information processing. In J.E. Desmedt (Ed.), Progress in Clinical Neurophysiology. Vol. 6: Cognitive Components in Cerebral Event-Related Potentials and Selective Attention. Basel: Karger, in press.

Chapman, R.M., McCrary, J.W., Chapman, J.A., & Bragdon, H.R. Brain responses related to semantic meaning. Brain and Language, in press.

Dixon, W.J. (Ed.) BMDP Biomedical Computer Programs. Berkeley: Univ. of California Press, 1975.
Programs developed at Health Sciences Computing Facility, UCLA; sponsored by NIH Special Research Resources Grant RR-3.

Heise, D.R. Evaluation, potency, and activity scores for 1551 words: a merging of three published lists. Chapel Hill, N. C., Univ. of N. Carolina, December 1971.

Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Miron, M.S. & Osgood, C.E. Language behavior: The multivariate structure of qualification. In R. B. Cattell (Ed.) Handbook of Multivariate Experimental Psychology. Chicago: Rand-McNally & Co., 1966, 790-819.

Osgood, C.E. The nature and measurement of meaning. Psychological Bulletin, 1952, 49, 197-237.

Osgood, C.E. Semantic differential technique in the comparative study of cultures. Amer. Anthro., 1964, 66, 171-200.

Osgood, C.E. Exploration in semantic space: A personal diary. Journal of Social Issues, 1971, 27, 5-64.

Osgood, C.E., May, W.H., and Miron, M.S. Cross-Cultural Universals of Affective Meaning. Urbana, Ill.: University of Illinois Press, 1975.

Rummel, R.J. Applied Factor Analysis. Evanston: Northwestern Univ. Press, 1970.