BRAIN AND LANGUAGE 11, 319–339 (1980)

Behavioral and Neural Analyses of Connotative Meaning: Word Classes and Rating Scales

ROBERT M. CHAPMAN, JOHN W. MCCRARY, JOHN A. CHAPMAN, AND JANICE K. MARTIN

University of Rochester

Behavioral (semantic differential) and neural (Evoked Potentials, EPs) responses were related to connotative meaning. The approach was based on Osgood's semantic analyses and dimensions of Evaluation (E), Potency (P), and Activity (A). The experimental variables were (1) the semantic class of the stimulus word (E+, E−, P+, P−, A+, A−) and (2) the dimension of the semantic scale (E, P, A) which the subject used to rate the stimulus words. These variables were experimentally combined such that on each trial the subject used a designated semantic scale to judge a specified stimulus word while brain activity was recorded. Using multivariate analyses, the effects on the EPs of stimulus word class, scale dimension, and their interaction were analyzed. The EP effects of stimulus word class were similar whether the subjects were merely saying the words or rating the words on a variety of semantic scales. Different EPs were found for six word classes, three semantic scale dimensions, and the 18 groups formed by their combination. The success rates in EP identification of (1) word class and (2) scale dimension did not depend on whether these two kinds of semantic variables involved the same or different semantic dimensions. The two kinds of semantic effects in EPs were largely independent. The behavioral data supported Osgood's results and showed that our subjects were appropriately processing the semantic information. The common analyses of data from all subjects suggest the universality of the connotative EP effects across individuals. This parallels, at the neural level, the universality of the connotative dimensions found at the behavioral level by semantic differential ratings. The EP effects imply that the neural representation of meaning is similar in different individuals.

INTRODUCTION

"The semantic differential, and the theory behind it, has been a revolution in conceptual and operational approaches to meaning" (Paivio, 1973, discussing Osgood's work, Osgood & McGuigan, 1973). The lawfulness

Supported in part by NIH Research Grant No. 5 R01 EY01593 from the National Eye Institute, ONR Contract N00014-76-C-0185, NIH Grant No. 5 P30 EY01319, and MH-33560. The authors thank Henry R. Bragdon for his collaboration in this research. Send reprint requests to Dr. Robert M. Chapman, Center for Visual Science, University of Rochester, Rochester, NY 14627.

0093-934X/80/060319-21$02.00/0

with which people give semantic differential ratings of words on various bipolar semantic scales has been demonstrated repeatedly (Osgood, 1952; Osgood, May, & Miron, 1975). The lawfulness of these data together with Osgood's view of meaning as a simultaneous bundle of distinctive features has resulted in quantitative descriptions of connotative meaning by the application of multivariate analytical techniques. The quantitative work of Osgood and his associates has found that connotative meaning space can be reasonably spanned by three underlying dimensions: Evaluation, Potency, and Activity. For example, the word COWARD might have scores of −.5 on Evaluation, −.7 on Potency, and +.2 on Activity, which quantitatively capture the following connotative meanings: "Quite Bad," "Very Weak," and "Slightly Active" (Osgood & McGuigan, 1973).

Our connotative meaning research with brain responses (Evoked Potentials, EPs) has been based on the results of Osgood's analyses. EPs contain information about semantic meaning not dependent upon the particular word stimuli (Chapman, 1974b; Chapman, Bragdon, Chapman, & McCrary, 1977; Chapman, McCrary, Chapman, & Bragdon, 1978; Chapman, 1978a, 1978b, 1978c; Chapman, 1979; Begleiter, Porjesz, & Garozzo, 1979). Combinations of components of these EPs were reliable indicators of semantic differences. In these experiments semantic meaning was manipulated by careful selection of stimulus words. In addition to internalized representations of connotative meaning elicited by stimulus words, another aspect of internalized representation may relate to an individual's semantic processing. When the same word is presented on different occasions, a subject may have different semantic expectancies and the semantic information in the words may be processed along different semantic dimensions. For example, a person might be primarily concerned with Potency (powerful–powerless) when a stimulus word NEWSPAPERS occurs or he might be primarily concerned with Evaluation (good–bad). Does the internal representation related to the word NEWSPAPERS vary when different semantic rating dimensions are being applied? Do these different semantic rating dimensions have their own internal representations?

In order to study questions of this sort, we manipulated the semantic rating dimensions by assigning various semantic differential scales to the subjects at different times. The subject's task was the semantic differential task 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 semantic dimensions were used (Table 1). Thus, basically a 6 × 3 factorial design (with some additional design features) was used: six semantic categories of words (representing opposite ends of the E, P, and A dimensions) combined with three kinds of semantic differential tasks (predisposing the subject for semantic processing along the E, P,

TABLE 1
LOADINGS OF SEMANTIC DIFFERENTIAL SCALES ON FIRST THREE FACTORS: EVALUATION (E), POTENCY (P), AND ACTIVITY (A)

Scale	Mean loadings of eight subjects		
	Factor		
	1	2	3
E Dominantly			
E1 nice–awful	.84	.14	−.04
E2 sweet–sour	.81	.04	−.04
E3 good–bad	.84	.19	.03
E4 heavenly–unheavenly	.81	.11	−.03
E5 mild–harsh	.78	−.16	−.09
Average	.82	.23	−.04
P Dominantly			
P1 big–little	.14	.68	−.01
P2 powerful–powerless	.15	.71	.20
P3 deep–shallow	.38	.46	−.09
P4 strong–weak	.29	.66	.16
P5 long–short	.17	.56	−.17
Average	.22	.62	.02
A Dominantly			
A1 fast–slow	−.03	.01	.62
A2 young–old	.30	−.32	.48
A3 noisy–quiet	−.25	.31	.49
A4 alive–dead	.46	.23	.44
A5 known–unknown	.36	.12	.22
Average	−.03	.06	.46

or A dimension). This permitted assessing the effect of the connotative meaning evoked by the words, the effect of the semantic judgment dimension induced by the semantic differential task, and their interaction.

Individual analyses of 10 subjects indicated that EP data from each of the subjects could be used to successfully discriminate among semantic word groups and among semantic scale dimensions (Chapman, 1979). The success rates varied little among EPs of the 10 subjects and lent further support to the ubiquitous nature of semantic effects in EPs.

In this paper, group analyses of the pooled EP data of all subjects are given, semantic word class effects are compared for "Say Word" and "Rate Word" tasks, the interaction of stimulus word class and rating scale dimension is investigated, and behavioral semantic-differential ratings are analyzed.

MATERIALS AND METHODS

The word stimuli were contained in two lists used in previous experiments (Chapman et al., 1977, 1978). These were compiled from the available E, P, and A glossaries (Osgood, Note 1; Heise, 1971). The words were relatively "pure" in the sense that they scored extremely positive or negative on one of the semantic dimensions and were relatively neutral on the other two. Twenty words from each of six semantic classes (E+, E−, P+, P−, A+, A−) were randomly assigned to a list. The words within each list were given in different random orders from run to run to prevent subjects anticipating either a semantic class or a particular word.

Five rating scales were selected to represent each of Osgood's three semantic dimensions (Evaluation, Potency, and Activity; Osgood, 1964). Each of the 15 semantic scales (Table 1) was used with each stimulus word.

Before each run the subject was assigned a rating scale, e.g., NICE–AWFUL, to be used on all 120 words in that run. The subject was asked to report a rating for each stimulus word on the designated semantic scale using values from +3 to −3. For example, if the rating scale was NICE–AWFUL and the meaning of the word was more NICE than AWFUL, a "+1," "+2," or "+3" rating was given to express the degree of NICE-ness. If the word was perfectly neutral on that scale, a "zero" was given. More detailed instructions may be found in Chapman (1979). These ratings constituted the data for the behavioral semantic analyses.

The subject sat in a dark, sound-damped chamber. Each word was individually presented as a briefly flashed stimulus on a computer-controlled CRT. The average word subtended a visual angle of 1.5° with a duration of 17 msec. 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), and (4) Blackout for 2.5 sec, allowing the subject time to report his semantic judgment of the word on the designated scale. A number of words (120) were presented in this fashion to constitute an experimental run.

During experimental runs, the subject's EEG was recorded from standard Grass electrodes (silver cup shape), attached to the scalp by bentonite CaCl paste. Data reported here were from a midline, central-parietal 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 by EOG (electrooculogram).

The 10 unselected subjects were six female and four male paid volunteers with a median age of 21 (range: 18–50). Over a series of sessions each of the subjects was given 30 runs of 120 words (20 words in each of six semantic classes). 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 semantic differential scales were given in different random orders for each subject.

The data were analyzed using multivariate procedures. The set of EPs for each subject was first standardized in a fashion similar to that reported by Chapman et al. (1978). Next, the standardized data were measured by Principal Components Analysis (Chapman, 1974a; Chapman et al., 1979). The EP component scores were used in Multiple Group Discriminant Analyses in order to relate the brain responses to various semantic groups (Chapman et al., 1978). The classification success rates were assessed by Chi-square tests. For each Discriminant Analysis, several kinds of classification success (EPs correctly identified) were evaluated: development, jackknifed cross-validation, and other-list cross-validation. The classifications of the cases (EPs) used to develop the classification functions are called development. The classifications of each case (EP) by classification functions based on the remaining cases (EPs) are called jackknifed cross-validation. The classifications of List 1 cases (EPs) by classification functions based on List 2 (EPs), and vice versa, are called other-list cross-validation. The details of these analyses are given below with the appropriate results.

RESULTS

(I) Behavioral Semantic Ratings

Over the course of the experiment, each subject rated every stimulus word on every semantic scale. The numerical ratings were available for eight of the subjects. Out of a maximum total of 28,800 ratings, 27,128 were available for analysis.

These ratings were analyzed to provide additional assessments of the semantic composition of the rating scales, the semantic properties of the stimulus word classes, and the agreement between the subjects' ratings of the stimulus words and the values in the glossaries from which the words were selected.

Separate Principal Components Analyses (BMDP4M, Factor Analysis Program; Dixon, 1975) were computed for the behavioral data of each subject. The input data matrix consisted of two lists of 120 stimulus words (240 cases) each of which was rated on 15 scales (15 variables). A complete data matrix for each subject contained 3600 ratings; the data matrix available for most subjects was slightly smaller (< 14 percent loss) due to missing data. Three factors were extracted from the 15 × 15 matrix of correlations among the ratings on the scales. The factors were rotated using the normalized varimax criterion. The three factors accounted for 49 to 79% (mean, 60%) of the total variance in the standardized ratings of the subjects. Factor scores were computed for each word on each of the three underlying dimensions.

The rotated factor loadings obtained from the individual analyses were averaged across subjects for each scale (Table 1). Since the factors were orthogonal, the loadings were also the correlations of the variables (rating scales) with the factors (semantic dimensions). Fisher's Z transform was used in obtaining the mean loadings (correlations). The rating scales are grouped according to the semantic dimension they were selected to represent on the basis of Osgood's (1964) results. The overall pattern of loadings is remarkably similar to Osgood's. The interpretation of the three dimensions is simplified by noting that the five highest loadings for Factor 1 are those of scales E1 to E5, the five highest for Factor 2 are P1 to P5, and the five highest for Factor 3 are A1 to A5. The three most important dimensions underlying the subjects' use of the scales appear to be the E, P, and A dimensions identified by Osgood.

The dominant semantic dimension of most of these scales is clearly defined, since the scales correlated highly with only one factor (Table 1). For example, the scale NICE–AWFUL has a high correlation (+.84) with Factor 1 (Evaluation) and low correlations (+.14, −.04) with Factors 2 and 3 (Potency, Activity). As in Osgood's reported loadings (1964), "pure" Activity scales are difficult to find. His factor results show similar tendencies for scales A2, A3, and A4 to load on the E dimension. The average loadings for the scale dimensions were obtained by averaging across the 5 individual scales. These averages indicate that, as a cluster of scales, the semantic composition of each scale dimension is quite well defined and distinct.

Factor scores for each word quantitatively measure the location of that word along each of the three underlying semantic dimensions. Table 2 summarizes for our subjects the mean factor scores obtained for each of the word classes previously formed by selecting words on the basis of the E, P, and A values in Heise's atlas (1971). This provides an operational definition of the meaning of the word classes in the eight-subject group. The meanings in terms of the factor scores agree with the intent of the selection procedure in forming the word classes. For example, the mean of the E+ words is clearly positive on the Evaluation dimension (+.67) and is nearly neutral on the Potency (+.11), and Activity (.00) dimensions.

The subjects' semantic judgments were directly compared to the Heise–Osgood normative data on a word by word basis. The degree of similarity was assessed by correlating our subjects' factor scores for the stimulus words with the E, P, and A values in Heise's atlas (1971). The averages and standard deviations of these correlation coefficients for the eight subjects are shown in Fig. 1. Factor 1 scores were most positively correlated with Heise–Osgood E values, Factor 2 scores with P values, and Factor 3 scores with A values. All other correlations were appropriately small. Thus, our subjects individually exhibited connotative semantic structures similar to those found in previous groups. All of the analyses of the behavioral data show that the subjects were rating the stimulus words in a semantically coherent way.

(II) Brain Responses (EPs)

Semantic word class effects when tasks are to say or to rate stimulus words. A previous experiment investigated EP effects associated with the same six semantic classes of words when the subject's task was merely to

TABLE 2
MEAN FACTOR SCORES OF WORDS WITHIN SEMANTIC CLASSES (EIGHT SUBJECTS)

	Word class					
	E+	E−	P+	P−	A+	A−
Factor 1 (Evaluation)	+0.67	−1.42	−.04	.25	.24	.19
Factor 2 (Potency)	.11	−.12	+0.81	−0.99	.16	.03
Factor 3 (Activity)	.00	.08	.01	−.09	+0.76	−0.75

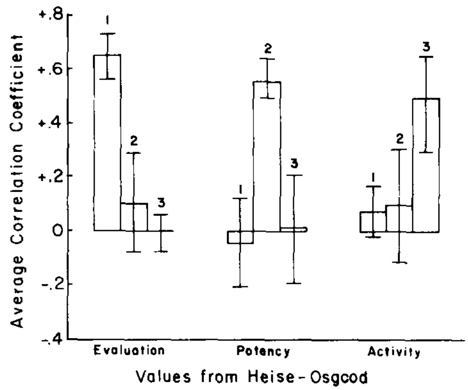

FIG. 1. Correlations of scores on Factors 1, 2, and 3 from subjects' ratings with Heise–Osgood values on Evaluation, Potency, and Activity semantic dimensions (Heise, 1971). Factors obtained from a separate Principal Components Analysis on each of eight subject's semantic differential ratings of stimulus words on 15 semantic scales. Factor scores for the stimulus words were correlated (Pearson) with the Heise–Osgood values (average $N = 228$); correlation coefficients were averaged for the eight subjects (computations used Fisher's r-to-z transform); mean $\pm$ 1 standard deviation shown.

say each word after it was flashed (Chapman et al., 1978). In the present experiment the subject's task was to give semantic differential ratings of each word on scales predominantly loaded on one of three semantic dimensions. Does the increased task complexity interfere with discriminating the word class by brain response measures? Does the use of different scales, loaded on different semantic dimensions, interfere with EP identification of the connotative class of stimulus words? Do the various semantic expectancies engendered by prior assignment of semantic scales interfere with identifying the stimulus word classes? Comparisons of the present experiment and previously reported results provide evidence for the generality and robustness of classifying stimulus word classes.

In order to facilitate comparisons with previous results, the data were analyzed in this section in as comparable a fashion as possible. This entailed collapsing the 15 rating scales by averaging the EPs across them. For each subject this resulted in separate EPs for six semantic classes each for Lists 1 and 2. The EP data were standardized separately for each subject (values at each time point brought to mean = 0 and SD = 1 for the set of EPs from each subject).

The average standardized EPs for the six semantic classes are superimposed for "Say Word" and "Rate Word" tasks in Fig. 2. The waveforms for these different tasks tend to be similar within each semantic word class and different between the six semantic word classes.

Following the procedure used with the "Say Word" data (Chapman et al.,1978), the standardized EPs for the "Rate Word" data for each subject

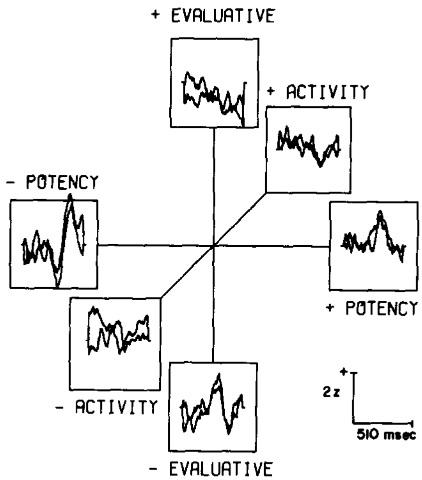

FIG. 2. Standardized brain responses for six semantic classes superimposed for "Say Word" and "Rate Word" tasks. Osgood's three-dimensional semantic space is represented schematically. EPs cover 510 msec (102 time points × 5 msec) beginning at the time the words were flashed. Each subject's data at each time point were transformed to z scores. Averages for each task contain data for two lists and two groups of 10 subjects (Chapman et al., 1978; Chapman, 1979). Monopolar recordings (bandpass: 0.1 to 70 Hz) from a scalp location ⅓ of the distance from Cz to Pz; positive is up relative to linked ear lobes.

were adjoined to form an input matrix of 120 EPs by 102 time points for a varimaxed Principal Components Analysis (BMDP4M; 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 EP measures entered into Discriminant Analyses (BMDP7M, Stepwise Discriminant Analysis Program; Dixon, 1975).

Discriminant classification analyses using the EP component scores to distinguish semantic word classes were of two kinds: (1) two word classes at a time and (2) six word classes at once. For both kinds, separate analyses were made of List 1 and List 2 data. For each analysis, cross-validations were obtained by the jackknifed procedure (one-left-out procedure) and by the cross-list procedure (applying the classification functions to the EP data obtained with the word list whose data were not used to compute the functions).

The results of classification analyses for two word classes at a time on the three semantic dimensions are given in Table 3. In each of the analyses, classification functions were computed which detected differences between the semantic groups (significant by Chi-square tests of classification success). 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

TABLE 3

CLASSIFICATION ANALYSES FOR TWO POLAR WORD CLASSES AT A TIME ON THREE SEMANTIC DIMENSIONS

Percentages of EPs correctly classified
Subject task: Semantic differential ratings.
(Subject task: say word)

Semantic dimension	Polarity	Development		Jackknifed cross-validation		Other-list cross-validation	
		List 1		List 1		List 2	
Evaluation	Positive	90	(100)[a]	80	(100)	90	(80)
	Negative	90	(100)	90	(100)	70	(80)
Potency	Positive	100	(100)	100	(100)	90	(100)
	Negative	100	(100)	90	(100)	80	(90)
Activity	Positive	90	(100)	90	(90)	50	(40)
	Negative	70	(90)	60	(80)	60	(50)
Overall		90.0	(98.3)	85.0	(95.0)	73.3	(73.3)
		List 2		List 2		List 1	
Evaluation	Positive	100	(100)	90	(70)	60	(80)
	Negative	90	(80)	90	(70)	80	(50)
Potency	Positive	100	(100)	100	(90)	90	(100)
	Negative	100	(90)	90	(90)	90	(90)
Activity	Positive	100	(100)	70	(100)	40	(100)
	Negative	100	(100)	90	(90)	90	(20)
Overall		98.3	(95.0)	88.3	(85.0)	75.0	(73.3)
Combined results		94.2	(96.7)	86.7	(90.0)	74.2	(73.3)

Note. Each individual percentage based on 10 EPs. Percentage correct expected by chance: 50%.

[a] Percentages in parentheses are from Chapman et al. (1978).

the table. In this case, the classification function classified EPs to E+ and E− classes with 90% accuracy. The jackknifed cross-validation (one-left-out procedure) success was 80 and 90% for E+ and E− classes. When the classification function developed from List 1 data were applied to List 2 data, 90% of the EPs obtained to E+ words were correctly assigned to the E+ class and 70% 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 these analyses.

The success in discriminating stimulus word classes along the Potency semantic dimension (P+ vs. P−) was also high. The accuracy of classifying EPs along the Activity semantic dimension (A+ vs. A−) was not as high as for the E and P semantic dimensions.

When the data for List 2 were used to develop the classification functions (bottom half of Table 3), the results in general were quite similar to those obtained when the development was based on List 1 data.

Overall, the analyses of two word classes at a time had an average development 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 10 subjects. When the same classification functions obtained using EP data from one word list were applied to the EP data obtained from the other word list, the overall success rate was 74%. These success rates are very similar to those obtained in previous results when the subjects were merely saying the words (compare 94 to 97%, 87 to 90%, 74 to 73%).

The results of classification analyses for all six word classes at once are given in Table 4. Again, classification functions were obtained which detected statistically significant differences among the groups. Since there were six semantic classes, the probability of correct classifications by chance is 1/6 or 16.7%. Separate Discriminant Analyses were developed for List 1 and List 2 data (upper and lower halves of Table 4). The development classification success rates (column 1) of assigning EPs elicited by words of particular semantic classes to the correct semantic classes were well above the chance level.

The overall jackknifed cross-validation success rates for six word classes were 43% for List 1 and 57% 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 57 and 45%, figures well beyond the chance level of 16.7%. This is a particularly stringent test of cross-validation, since it assesses the ability to generalize to a different list of words (only P− words being the same), as well as to generalize to individual data from 10 different subjects.

The success rates of identifying word classes obtained with EP data collected while the subjects were doing semantic differential ratings

TABLE 4
CLASSIFICATION ANALYSES FOR SIX WORD CLASSES AT ONCE

Percentages of EPs correctly classified
Subject task: Semantic differential ratings.
(Subject task: say word)

Semantic dimension	Polarity	Development List 1		Jackknifed cross-validation List 1		Other-list cross-validation List 2	
Evaluation	Positive	60	(30)[a]	40	(30)	70	(50)
	Negative	50	(80)	30	(50)	60	(50)
Potency	Positive	70	(40)	60	(40)	70	(30)
	Negative	100	(80)	80	(80)	80	(70)
Activity	Positive	50	(50)	10	(20)	20	(10)
	Negative	40	(60)	40	(30)	40	(30)
Overall		61.7	(56.7)	43.3	(41.7)	56.7	(40.0)
		List 2		List 2		List 1	
Evaluation	Positive	60	(60)	50	(60)	40	(70)
	Negative	70	(70)	70	(60)	60	(50)
Potency	Positive	60	(30)	40	(20)	20	(20)
	Negative	80	(90)	70	(80)	70	(70)
Activity	Positive	40	(20)	30	(0)	50	(20)
	Negative	80	(50)	80	(40)	30	(10)
Overall		65.0	(53.3)	56.7	(43.3)	45.0	(40.0)
Combined results		63.4	(55.0)	50.0	(42.5)	50.9	(40.0)

Note. Each individual percentage based on 10 EPs. Percentage correct expected by chance: 16.7%.
[a] Percentages in parentheses are from Chapman et al. (1978).

(Table 4) compare favorably with data obtained while subjects were saying the word (Table 4, values in parentheses). The jackknifed cross-validation rates, combining results for List 1 and List 2, are compared for "Say Word" and "Rate Word" experiments in Fig. 3. The level and pattern of success rates appear quite similar. Increased task complexity had no effect on the success in Discriminating word classes. In general, the semantic classes of the stimulus words were identified from EP data as successfully when the subjects were rating the stimulus words as when simply saying the words.

Brain response effects for semantic class of stimulus words and semantic dimension of rating scales. The experimental design made it possible to examine two additional, key questions: (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 according to whether the stimulus word and the task scale involve the same or different semantic dimensions)?

For the purpose of answering these questions, the EP data of each of the subjects were averaged to represent each of the 36 experimental combinations of 6 semantic classes of stimulus words × 2 lists of words × 3 rating scale dimensions. These EPs as a set were standardized at each time point (Mean = 0, SD = 1) separately for each subject. Average EPs before and after standardizing are shown in Figs. 4 and 5, respectively.

Varimaxed principal components analysis of the combined data of all 10 subjects (360 EPs × 102 time points) resulted in retaining 11 components accounting for 93.4% of the total variance. The scores for these components were used as the EP measures entered into discriminant analyses.

Classification functions were computed to distinguish among all 18 semantic conditions defined by the six semantic classes of words in

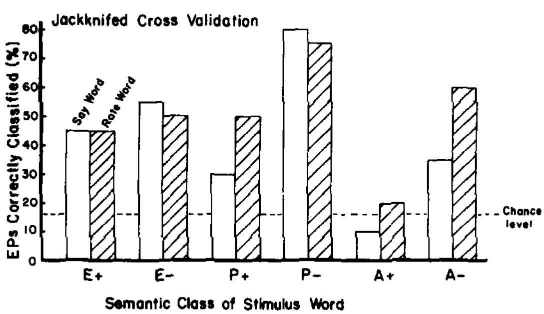

FIG. 3. Classification success for six word classes compared for "Say Word" and "Rate Word" tasks.

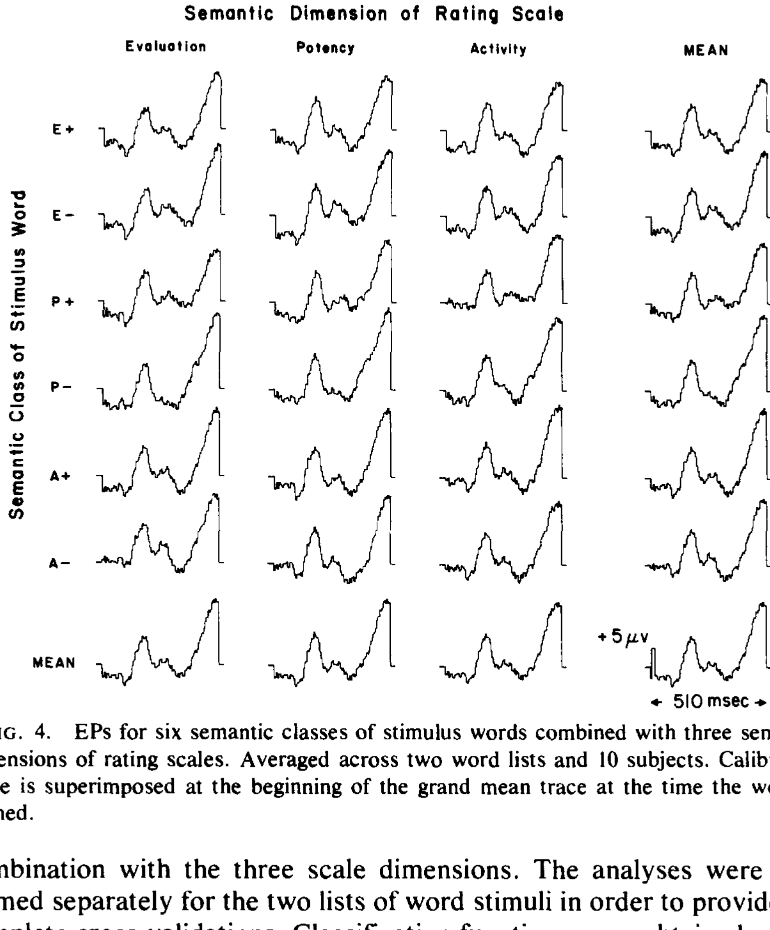

FIG. 4. EPs for six semantic classes of stimulus words combined with three semantic dimensions of rating scales. Averaged across two word lists and 10 subjects. Calibration pulse is superimposed at the beginning of the grand mean trace at the time the word is flashed.

combination with the three scale dimensions. The analyses were performed separately for the two lists of word stimuli in order to provide for complete cross-validations. Classification functions were obtained which detected statistically significant differences among the groups. 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 5.

Since there are 18 groups to which an EP could be assigned, one out of 18 or 5.6% of the EPs would be expected to be correctly assigned by chance. The average development classification success rate obtained when classifying the EPs used to develop the functions was 28%: five 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 14%, it remains 2.5 times better than chance.

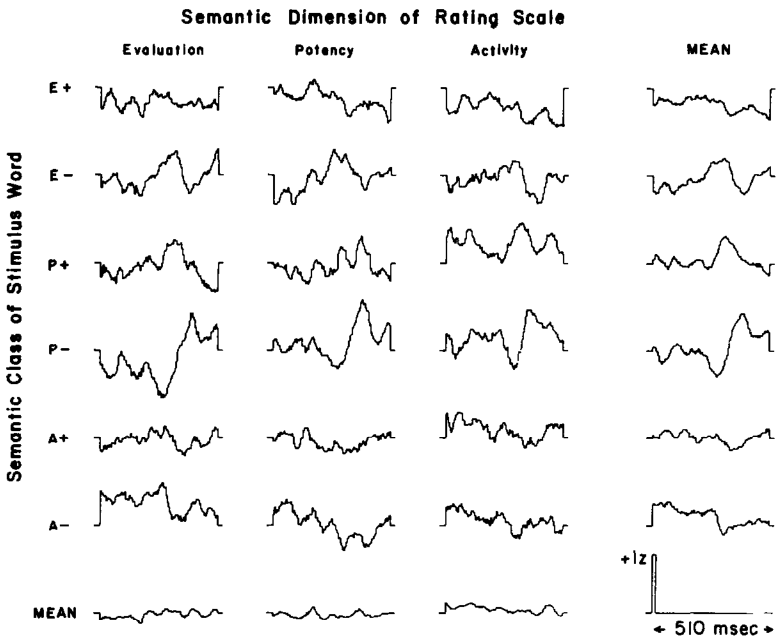

FIG. 5. Standardized EPs for six semantic classes of stimulus words combined with three semantic dimensions of rating scales. Same as Fig. 4 data except each subject's data at each of the 102 time points is standardized (transformed to z scores). As demonstrated here, the grand mean standardized potential is a flat line at zero; calibration pulse is superimposed at the time the word is flashed. Differences among standardized mean EPs in margins show word class effects (right column) and rating scale effects (bottom row).

TABLE 5
18 CLASSES: SIX WORD CLASSES BY THREE SCALE DIMENSIONS

Word class	Percentage of EPs correctly classified											
	Development				Jackknifed cross-validation				Other-list cross-validation			
	E	P	A	Mean	E	P	A	Mean	E	P	A	Mean
E+	20	30	30	27	10	20	30	20	20	10	30	20
E−	40	25	25	30	20	15	5	13	20	5	0	8
P+	20	25	35	27	5	15	10	10	5	10	15	10
P−	50	55	45	50	30	30	20	27	30	10	0	13
A+	10	5	20	12	0	0	10	3	10	0	5	5
A−	40	10	15	22	20	10	5	12	15	5	10	10
Mean	30	25	28	28	14	15	13	14	17	7	10	11
χ^2 (1 *df*)				334.6				46.1				20.1

Note. E = Evaluation, P = Potency, A = Activity Semantic Dimension. Results combined for two word lists; 10-subject group. Each individual percentage based on 20 EPs. Percentage correct expected by chance: 5.6%.

The third part of Table 5 (Other-List Cross-Validation) presents the results obtained when the classification functions were applied to data not used in their development and collected using a different list of words. As might be expected the overall success rate is lowered. However, the 11% accuracy is nearly double the percentage correct expected by chance and the chi-square test supports this difference as statistically reliable ($p <$ .001).

The table reveals considerable variability in accuracy with which the 18 combinations are identified. Those combinations which involved the A+ class of words were detected less accurately than others throughout all evaluations. The other-list cross-validations remain impressive only when semantic judgments about the word classes involve scales representing the evaluative dimension.

The use of the semantic differential task in conjunction with the six categories of words would be expected to predispose the subjects for semantic processing along the E, P, or A dimension. This could result in an interaction of semantic effects which would be represented in the EPs and influence the outcome of classifications. The 18 group Discriminant Analyses and classifications of the EPs enable us to examine this question of the interrelationship of word dimensions and scale dimensions.

Table 6 presents a simple retabulation of the development group data. As in Table 5, an EP was counted as correctly classified only if both the word class and scale dimension were appropriately tagged. Partially correct was counted as incorrect. In Table 6, the percentages of such correctly identified EPs are presented according to whether the semantic dimensions of the word class and of the scale coincided or differed. Chi-square tests of independence were computed for these retabulations. The number of correct classifications into word classes is independent of whether the rating scale is of the same or a different semantic dimension (Chi-square (5 *df*) = 3.72, $p >$ .50). Similarly, the correct identification of the rating scale dimension is not affected by the word class being of the same or different dimension (Chi-square (2 *df*) = .01, $p >$ .99).

These analyses of classification data indicate that, as represented in the EP, the semantic processing of word stimuli and the processing imposed by a semantic rating task do not become entangled. They do not interact in such a way as to influence (enhance or suppress) the detectability of one another.

Since the analyses of the classifications above indicate that the effects in EPs related to distinguishing word classes are independent of distinguishing rating 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

TABLE 6
COMPARISONS OF CLASSIFICATIONS WHEN THE SEMANTIC DIMENSION OF WORD CLASS AND SCALE ARE THE SAME OR DIFFERENT

Percentage of correct classifications (development)							
	Word class						
Relation of scale dimension to word class	E+	E−	P+	P−	A+	A−	Average
Same	20	40	25	55	20	15	29
Different	30	25	28	48	8	25	27

Individual percentages based on 20 EPs for "Same" and 40 EPs for "Different"

	Scale dimension			
	Evaluation	Potency	Activity	Average
Same	30	40	18	29
Different	28	38	16	27

Individual percentages based on 40 EPs for "Same" and 80 EPs for "Different"

Note. Classification functions designed for 18 classes (six word classes by three scale dimensions). Results combined for two word lists; 10-subject group. Percentage correct expected by chance: 5.6%.

other kind of 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 (Tables 5 and 6). For the present purposes, however, the Discriminant Analyses were focused 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 7 and 8. For both kinds of analyses, separate Discriminant Analyses were made on the data obtained with the two word lists and the classification percentages averaged. For both Tables 7 and 8 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 list of words, respectively.

Separate identification of word class (Table 7) had an overall development success rate of 48%, which is to be compared with a chance rate of

TABLE 7
CLASSIFICATION FOR SIX WORD CLASSES, IGNORING RATING SCALES

Word classes	Percentages of EPs correctly classified		
	Development	Jackknifed cross-validation	Other-list cross-validation
E+	33	32	30
E−	50	43	42
P+	55	47	37
P−	75	73	67
A+	27	18	18
A−	48	35	25
Average	48.0	41.4	36.4
χ^2 (1 *df*)	253.1	156.6	99.4

Note. Results combined for two word lists: 10-subject group. Each individual percentage based on 60 EPs. Percentage correct expected by chance: 16.7%.

17% (six word classes). The generality of the classification functions is indicated by the jackknifed cross-validation success rate (41%) and other-list cross-validation success rate (36%). These analyses indicate that word classes can be successfully identified despite the wide variety of semantic scales used by the subjects when these data were obtained.

Separate identification of scale dimension (Table 8) had an average development success rate of 50%, 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-list cross-validation is weak (36%). These analyses indicate that semantic scale dimensions can be successfully identified within a given list despite the wide variety of words used as the specific stimuli for the EPs.

TABLE 8
CLASSIFICATION FOR THREE RATING SCALE DIMENSIONS, IGNORING STIMULUS WORD CLASSES

Scale dimension	Percentages of EPs correctly classified		
	Development	Jackknifed cross-validation	Other-list cross-validation
Evaluation	48	44	40
Potency	55	48	38
Activity	46	39	31
Average	50	44	36
χ^2 (1 *df*)	42.8	16.7	1.1

Note. Results combined for two word lists: 10-subject group. Each individual percentage based on 120 EPs. Percentage correct expected by chance: 33.3%.

The identifications of word classes and 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. The reasons for this may lie in the tertiary role that the A dimension plays in semantic-differential judgments. Osgood and others (Osgood et al., 1975) have 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 1 shows that for our subjects also the A scales have lower loadings on their dominant dimension and higher loadings on their nondominant dimensions than do the E or P scales. In a similar vein, the average score difference between the word classes on their respective dominant dimensions (Table 2) was only 1.5 between the A+ and A− word classes, whereas it was 2.1 between the E+ and E− classes, and 1.8 between the P+ and P− classes. These semantic quantifications derived from behavioral measurements are consonant with our classification rates derived from brain response measures.

In general, the separate identifications of word classes and scale dimensions (Tables 7 and 8) were significantly better than chance, with the exception of the other-list cross-validation of the scale dimensions. It is to be noted that these success rates were obtained across subjects, i.e., the same classification functions were successfully used for all 10 subjects.

DISCUSSION AND CONCLUSIONS

The data presented in this paper, and in previous reports (Chapman, 1974b; Chapman et al., 1977, 1978; Chapman, 1978a, 1978b, 1978c, 1979), show that representations of connotative meaning of stimulus words exist in brain activity recorded from the scalp. The specificity of the EP effects is indicated by their relationship to semantic variables and lack of relationship to other kinds of variables. Distinctive brain response effects have been found for a number of semantic distinctions, including six semantic classes, three semantic scale dimensions, and the 18 groups formed by their factorial combination. The case for specificity of semantic effects is strengthened by the relatively fine-grain mapping of brain activity onto semantic space. If only two semantic classes had been used, then only one distinction could be made and the risk would be increased of confounding the semantic distinction with nonsemantic variables, such as attention, stimulus intensity, depth of processing, etc. The evidence against various nonsemantic alternative explanations has been given elsewhere (Chapman et al., 1978; Chapman, 1979). If only one EP component were being affected to varying degrees, it is unlikely that more than two semantic classes could be discriminated by the EP data. In the present analyses, a number of independent EP components were found to be important to the semantic discriminations. The orthogonality of the EP components is one of the conditions imposed by the Principal Components Analyses used to measure the EPs.

Different EP components tended to be important in discriminating (1) among the semantic word classes and (2) among the rating scale dimensions. This agrees with the previous analyses of the individual subjects (Chapman, 1979). In the group analyses reported here, the importance of various EP components for discriminating the semantic conditions was assessed by the magnitudes of the F values at step 0 of the Discriminant Analyses (BMDP7M). There was good agreement ($r = +.76$) between the Fs for discriminating word classes for List 1 and List 2 data, indicating that the same EP components were important. The Fs for discriminating word classes did not agree with the Fs for discriminating rating scale dimensions (rs $= -.32, +.01, -.15, -.04$), indicating that the same EP components were not important for word classes and rating dimensions. However, the Fs for discriminating rating scale dimensions were not consistent for the two lists ($r = -.19$). The latter result may be related to the small size of the rating scale effect compared to the robust word class effect.

The finding that different EP components may be related to the semantic class of the stimulus word and the semantic dimension of the rating scale points to an interesting difference from Osgood's analyses. Osgood's analyses relate both the stimulus words and the rating scales to the same underlying semantic dimensions (E, P, and A). The stimulus words are behaviorally measured by the ratings on the scales. At the behavioral level, the measurement is of the relation between the stimulus word and the rating dimension. At the neural level, however, it is possible to measure separately the effects of stimulus word and rating dimension. These two kinds of semantic variables need not relate to the same underlying neural dimensions. The evidence thus far indicates that different neural components are involved in the two kinds of semantic variables. The neural independence of the semantic class of the stimulus words and the semantic dimensions of the rating scales is further supported by the lack of interaction between these two kinds of semantic variables in the classification success rates (Tables 5 and 6). The classification rates were essentially unaltered when the word class and the rating scale belonged to the same or different semantic dimensions, e.g., the word "newspapers" when rated on a Potency or Evaluation scale.

Behavioral data were obtained simultaneously with the neural data. The analyses of the semantic differential ratings on 15 bipolar scales show that our subjects were appropriately processing the semantic information (Tables 1 and 2). Moreover, our subjects' ratings showed good agreement with the Heise–Osgood results (Fig. 1). These behavioral results support in detail Osgood's fundamental analysis of connotative meaning.

The finding that the EP effects related to connotative meaning of stimulus words hold across subjects suggested that the neurophysiological representation of meaning may be similar in different individuals (Chapman et al., 1978). This finding has been substantiated here (Tables 3, 4, 5, 6, Fig. 3). Furthermore, the rating scale dimensions could be identified not only by separate analyses of each individual's EP data (Chapman, 1979), but also by a set of classification functions developed for the entire group of subjects (Tables 5 and 8). The success of these analyses supports the universality of the connotative EP effects across individuals, which parallels, at the neural level, the universality of the connotative dimensions found at the behavioral level by semantic differential ratings (Osgood et al., 1975; Tables 1 & 2, Fig. 1 here).

Are the semantic EP effects related to stimuli, to responses, or to mediational processes? If the EP effects found between semantic word classes were related to the response end of internal processing, then the subject's task might be expected to alter the EP effects. Data were compared when the subject's tasks were quite different, namely simply saying the word vs. rating the word on a semantic scale. With the considerably different cognitive operations imposed by these tasks, essentially the same EP effects of stimulus word class were found. The EP waveforms for the six semantic word classes were similar (Fig. 2) and the classification success rates were similar (Tables 3 and 4, Fig. 3). The possibility that differences in the average EPs for the various semantic word classes might be attributable to consistent physical differences in the stimulus words is guarded against by the use of a relatively large number of words within each semantic class. Statistical analyses of the total luminance flux and luminance patterns in the stimulus words in the various semantic classes were at chance levels (Chapman et al., 1978). Since the word class EP effects do not appear to be tied to particular stimulus parameters and are not altered by varying the response end of internal processing, it seems reasonable to relate the EP effects to mediational processes. Thus, the EP effects related to semantic word class may be mediational neural processes related to bundles of distinctive features encoding connotative meaning.

REFERENCES

Begleiter, H., Porjesz, B., & Garozzo, R. 1979. Visual evoked potentials and affective ratings of semantic stimuli. In H. Begleiter (Ed.), *Evoked brain potentials and behavior*. New York: Plenum Press. Pp. 127–141.

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*, precirculated 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*, precirculated abstracts. Brussels: Presses Universitaires de Bruxelles. Pp. 43–45. (b)

Chapman, R. M. 1978. Language and evoked potentials. In D.A. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research*. Washington, D.C.: U.S. Gov't. Printing Office, EPA-600/9-77-043. Pp. 245–249. (a)

Chapman, R. M. 1978. Method of EP analysis in linguistic research. In D.A. Otto (Ed.), *Multidisciplinary perspective in event related brain potential research.* Washington, D.C.: U.S. Gov't. Printing Office EPA-600/9-77-043. Pp. 265–266. (b)

Chapman, R. M. 1978. On the road to specific information in evoked potentials. In J.C. Armington, J. Krauskopf, & B.R. Wooten (Eds.), *Visual psychophysics and physiology*. New York: Academic Press. Pp. 469–480. (c)

Chapman, R. M. 1979. Connotative meaning and averaged evoked potentials. In H. Begleiter (Ed.), *Evoked brain potentials and behavior*. New York: Plenum Press. Pp. 171–196.

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. 1979. 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. Pp. 80–105.

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

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

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

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

Osgood, C. E. 1964. Semantic differential technique in the comparative study of cultures. *American Anthropology*, 66, 171–200.

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

Osgood, C. E., & McGuigan, F. J. 1973. Psychophysiological correlates of meaning: Essences or tracers? In F.J. McGuigan & R.A. Schoonover (Eds.), *The psychophysiology of thinking*. New York: Academic Press. Pp. 449–483.

Paivio, A. The discussion of Dr. Osgood's paper. 1973. In F.J. McGuigan & R.A. Schoonover (Eds.), *The psychophysiology of thinking*. New York: Academic Press. Pp. 483–492.

REFERENCE NOTE

1. Osgood, C.E. Personal communication.