Memory & Cognition
1978, Vol. 6 (5), 481-490

Categorizing with overlapping categories

HIRAM H. BROWNELL and ALFONSO CARAMAZZA
Johns Hopkins University, Baltimore, Maryland 21218

An important aspect of the categorization process is that an item can be assigned membership in more than one semantic category. Previous work examining how subjects decide an item's membership in one of several alternative categories has most often used categories having a strict hierarchical relationship (e.g., bird-canary). Four experiments are reported that examine how subjects decide membership of simple pictorial stimuli in partially overlapping categories (e.g., high and very high). Experiment 1 was a rating task designed to identify items as members, nonmembers, or as falling on the fringes for several overlapping categories. In Experiments 2-4, this information was used to predict subjects' mean reaction time in speeded categorization tasks using the same pictorial stimuli. Subjects interpreted the categories in one of two very different ways. According to the first interpretation, there was a strict set-subset relationship between categories such as "high" and "very high." According to the second, the entailment relationship did not hold; membership in the category "very high" did not imply membership in "high." Even when subjects used a set-subset interpretation of the category labels, their reaction times were affected by a form of semantic response competition. Subjects took longer to verify an item's membership in a category when there was another more appropriate category descriptor for that item included in the experiment.

Two central features of the process of categorization are (1) that a single label can be applied to a set of items that need not be identical but that presumably share important common elements, and (2) that alternative labels can be applied to the same item. The first of these two aspects of categorization has received considerable attention (see Smith, in press, for a review). The second has been addressed primarily in research in which hierarchically related labels can be applied to the same object (e.g., Collins & Quillian, 1967; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). However, there has been little if any research that has addressed the question of how subjects proceed in categorizing objects that have several potential labels that are not hierarchically related (i.e., as in the case where one might describe a house as large or very large, etc.).

There are several different levels of specificity that a speaker may apply to a given stimulus. Conferring membership in a relatively restricted category gives more information about an object than membership in a more general category, and the level of description used to describe an object in conversation will vary as a function of the situation (Grice, 1975). One common and effective way to adjust the amount of information conveyed about an object by its label is through the use of adverbial modifiers, such as "very," "extremely," or "sort of." A ceiling can be labeled simply as high, or, alternatively, as very or extremely high. The use of intensifiers (e.g., very, extremely) restricts the meaning of the base concept (i.e., ceiling). There will be some ceilings that can be adequately described as high, but not as very or extremely high. A deintensifier (e.g., sort of) expands the extension of the base term (Lakoff, 1973). There may be some ceilings that can be described as "sort of high" but not as "high." The specific question addressed in the research reported below concerns the factors that affect the categorization of a pictorial stimulus when the stimulus can logically and correctly be assigned membership in any of several different categories.

This issue was addressed within the context of dimensional adjectives (e.g., high, low, etc.). Dimensional adjectives were used because membership in these categories depends only on the degree to which a pictorial stimulus possesses a single quality or property (e.g., height) that can be easily manipulated experimentally. Another reason for using dimensional adjective category concepts is that they are relative or context-dependent concepts. Given a set of values on a dimension such as size, some objects can be labeled large, others small, others very large, others sort of small, and so on.

There is some precedent for using dimensional adjectives to investigate problems in psychological semantics. Hersh and Caramazza (1976) have examined the modification of the base concepts "large" and "small" by various adverbial modifiers (e.g., very, extremely, sort of, etc.). The important point about this work is that it moves beyond the earlier attempts in psychology to quantify meaning of adjectives as points on a scale (e.g., Osgood, 1952) and provides

This research was supported by Biomedical Support Grant 5 S07 RR 07041-11 to the Johns Hopkins University. We would like to thank Rita Berndt and Harry Hersh for their helpful comments on an earlier draft of this paper. We also thank Joan McDermott for her assistance in the preparation of stimuli and for testing subjects for Experiment 2, and Greg Harmon and William Laufer for their assistance in data analysis. Please address requests for reprints to Alfonso Caramazza.

a model for viewing the application of a term as defining a range of values along some dimension(s). Subjects in these experiments typically were shown category labels one at a time, and, for each label, they were presented with a series of squares of varying size. The task was to respond "yes" or "no" whether a phrase was an appropriate descriptor for each square size. (Before starting, subjects viewed all possible square sizes.) To describe their results, Hersh and Caramazza (1976) used a construct from fuzzy-set theory (see Zadeh, 1965) called "grade of membership." Briefly, each stimulus has a value specifying the extent to which it is assigned membership in a given category or class. The numerical values for grade of membership range from 1.0, indicating clear membership, to 0.0, indicating clear nonmembership. A value of .5 indicates maximum uncertainty. In the experiments that follow, as well as in those reported by Hersh and Caramazza (1976), grade of membership is defined empirically as the proportion of "yes" responses that a pictorial stimulus is a member of a given target category. For example, if 9 out of 10 subjects agree that a particular square size is very large, then that stimulus would have a grade of membership of .9 in the category "very large."

The experiments reported below were similar in design to those reported by Hersh and Caramazza (1976), except that speeded verification tasks (in addition to a rating task) were used to examine categorization. The categories "high" and "low" were used both alone and in combination with adverbial modifiers such as "very" and "sort of." The pictorial stimuli were simple: single black dots at any of 10 equally spaced heights inside outlined rectangles of a constant size (Figure 1). Four experiments were performed. Experiment 1 was a rating task designed to obtain grade-of-membership functions for the phrases "high," "low," "very high," and "very low." In addition to these four phrases, the experiment contained "extremely high," "extremely low," "fairly high," "fairly low," "sort of high," and "sort of low" to provide a rich experimental context of overlapping categories. In particular, the categories "high" and "low" were bordered on both sides by overlapping categories: "Extremely high/low" and "very high/low" represented more extreme values on the height dimension, and "fairly high/low" and "sort of high/low" represented less extreme values on the height dimension. Definitions for phrases consisted of a listing of the degree of membership for each dot position. On the basis of data from Experiment 1, it was possible to specify definitions for a category in that particular experimental context. Experiments 2-4 examined how subjects' categorization of the same pictorial stimuli under time pressure would change from one experimental context to another. In Experiment 2, the context was the same as in Experiment 1. Several overlapping categories were used in the design, such that a single stimulus could be assigned membership in more than one category. In Experiment 3, the number of categories was reduced to two, "high" and "low." In this context, the pictorial stimuli (with the exception of those stimuli that fell at the border between the two categories) could be assigned membership in only one category. The final experiment included four categories: "high," "low," "sort of high," and "sort of low." In this context, the categories "high" and "low" overlapped with "sort of high" and "sort of low," respectively, but they (high and low) did not overlap with any categories that were more extreme on the height dimension.

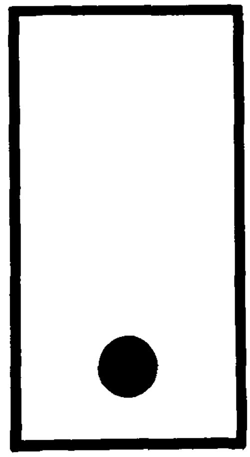

Figure 1. Pictorial stimulus showing Dot Position 2.

Our expectation was that reaction time (RT) would be affected by the degree of membership of an item in a category; that is, RTs for deciding membership for clear members should be faster than for those items that do not have such a privileged status. However, it was not clear how RT for deciding membership for an item in a category would be affected by the presence of an equally good alternative category.

EXPERIMENT 1

The first experiment used a rating task to obtain the extensional definitions for several adjacent categories labeled by the phrases "high," "very high," "low," and "very low." These four category labels were embedded in a rich context that included several other descriptors (e.g., extremely high, fairly low, etc.). The additional context phrases used in the experiment provided subjects with a full range of potential labels for the pictorial stimuli.

Method

Subjects viewed category labels paired with pictures of dots at different heights inside outlined rectangles (see Figure 1). They responded "yes" or "no" as to whether a phrase accurately described a dot's height or position in the rectangle and then gave a confidence rating for their responses.

Subjects. Thirteen students at the Johns Hopkins University participated in the experiment.

Stimuli. Slides of the four phrases (high, low, very high, and very low) were paired with slides of the 10 different dot positions. When presented on a screen, the rectangles subtended a visual angle of 6.0 by 10.6 deg. At the diameter, the dots subtended an angle of 1.5 deg, and the distance between adjacent dot positions measured from center to center subtended an angle of .9 deg. There were three replications of each of the 40 phrase-picture pairs, making a total of 120 trials. In addition to these experimental trials, slides of six other phrases (extremely high, extremely low, fairly high, fairly low, sort of high, and sort of low) were paired once with the 10 dot positions for an additional 60 trials. The total set of 180 trials was presented in a random order, using slide projectors equipped with electric shutters. On each trial, a phrase and a picture were presented simultaneously for 2.0 sec, with phrases on the left and pictures on the right side of a screen.

Procedure. Subjects were tested individually. A subject started each trial with a buttonpress, which caused a phrase-picture pair to appear for 2.0 sec. The subject responded orally "yes" or "no" as to whether the phrase applied to the dot's position in the rectangle and then gave a confidence rating on a scale from 1 (pure guessing) to 5 (complete certainty). The instructions stressed that the subject should use his own intuitions about language to make his decisions. The instructions also pointed out that a particular phrase could apply to more than one dot position, and that on some trials there would be no clearly right or clearly wrong answer. In addition, the subject was told to notify the experimenter if he changed his mind about a response. Before starting, the experimenter read the 10 phrases to the subject and showed him examples of all the dot positions.

Results

Subjects' responses were used to construct membership functions for the categories "high," "very high," "low," and "very low." The grade of membership for a dot position was defined simply as the proportion of "yes" responses that a phrase described a particular dot position. The confidence ratings were analyzed separately. The binary decisions (yes or no) and the confidence ratings were incorporated into a single scale that covered the same range (0.0-1.0) as the membership scale, based on proportions. The conversion formula used was that used by Hersh and Caramazza (1976): If d is the binary decision (1 = yes, −1 = no), and r is the value of the confidence scale ($1 \leq r \leq 5$), then the assigned value is defined as: scale value = $.5 + d(r/10)$. Finally, the average of the three replications presented to a subject for each phrase-picture pair was computed, and these means were collapsed across subjects. The function based on the derived scale values were then compared to those based on proportions of "yes" responses. The correlations between the two kinds of membership curves ranged from $r = +.97$ to $r = +1.00$. Based on the magnitude of these correlations, the following discussion is limited to the functions based on proportion of "yes" responses.

Initial examination of the data indicated that subjects interpreted the category labels in the experiment in one of two very different ways, just as subjects had in the Hersh and Caramazza (1976) study. One interpretation was what Hersh and Caramazza have called "logical." If, for example, a dot position was called "very high" or "extremely high," it was also called "high"; the extensions for the modified categories "very high" and "very low" formed proper subsets of the extensions for the base concepts "high" and "low." In other words, being very high entailed being high. Subjects who responded "yes" to the phrase "high" paired with Dot Position 1 (the top of the rectangle) and "yes" to "low " paired with Position 10 (the bottom) were placed in this "logically" responding group.

Another group of subjects responded as if the base categories were bounded: If a dot position was very high, it could not also be high, and vice versa. Unlike the "logical" group described above, subjects in this group did not interpret very high as entailing high. This kind of response will be referred to as a "linguistic" interpretation, again following the usage of Hersh and Caramazza (1976). The responses of these two groups of subjects represented qualitatively different interpretations of the terms' meanings, and for this reason data from the two groups of subjects were analyzed separately.

Results from the logically responding group, which included 10 of the 13 subjects, will be discussed first. The upper panel of Figure 2 shows the membership

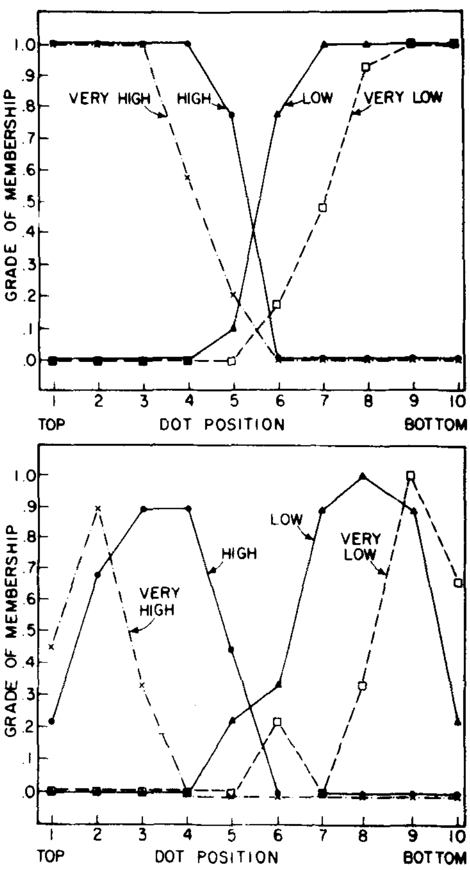

Figure 2. Membership functions for logically responding subjects (upper panel) and linguistically responding subjects (lower panel).

functions for "high," "very high," "low," and "very low" for these subjects. There are two things to notice in these data. One is that "very" acts as an intensifier on the base concept "high," in that it causes a shift toward the upper end of the rectangle; it restricts the extension of the base concept. The second aspect of these data is that for both phrases there is a spread of three to four dot positions where there is not complete agreement about what is and what is not a member of the category. The extension for the category "very high," for example, blends slowly into that for "high."

The membership curves for "low" and "very low" are quite similar to those for "high" and "very high." In fact, the curves for "low" and "very low" are almost symmetrical to those for "high" and "very high." The effect of "very" as a modifier was to restrict the extension of the base concept "low," and there was a spread of three to four dot positions over which subjects did not give consistent membership judgments.

The lower panel of Figure 2 shows the data for the linguistic group of subjects, which included 3 of the 13 subjects. A subject was placed in this group if he responded "no" to "high" at Position 1 (the top of the rectangle) or "no" to "low" at Position 10 (the bottom of the rectangle). Apparently, these subjects felt that a phrase applied to a dot position only if there were no more appropriate phrases available. In other words, if a particular dot position was best described by "very high," it was not adequately described by "high," and likewise for "very low" and "low." This interpretation of the data is consistent with subjects' reports. The functions suggest that "very high" and "very low" were in turn bounded by "extremely high" and "extremely low," two of the context phrases included in the experiment. In other respects, the curves for this group of subjects were similar to those for the logically responding subjects. "Very" produced a restriction of the base terms, making the functions for "very high" and "very low" more peaked; and "very" caused a shift in the base terms' extensions toward the extremes of the rectangle. Finally, the functions for "high" and "low" were roughly symmetrical.

EXPERIMENT 2

The purpose of the first experiment was to obtain membership functions or "definitions" for the phrases "high," "low," "very high," and "very low." In turn, these functions were used to specify the overlap among the different category concepts included in the experiment and to locate the boundaries of the categories. Several speeded categorization experiments were then performed to examine how subjects decided category membership for pictorial stimuli that potentially could be assigned membership in more than one category. The three RT studies differ in the composition of the stimuli. In Experiment 2, the stimulus set from Experiment 1 was used again. In this set, there were several overlapping categories. Although the category "high," for example, extended from roughly the midline of the rectangle to the top, there were categories in the experiment whose meanings were more extreme on the height dimension (e.g., very high) and categories whose meanings were less extreme (e.g., sort of high). Thus, subjects had to verify category membership for items having simultaneous membership in other categories.

Method

On each trial, subjects viewed a category label first, and then a picture of a dot inside a rectangle. While timed, they responded "true" or "false" whether a phrase accurately described a dot's position in the rectangle.

Subjects. Twenty Johns Hopkins students who had not participated in Experiment 1 served as subjects.

Procedure. The procedure from Experiment 1 was followed except as noted below. On each trial, a single phrase appeared for 2.0 sec, followed by a picture for 2.0 sec. The onset of the picture started a millisecond clock that ran until a subject's response (true or false) was spoken into a voice key. Before the start of the experiment, a subject received 10 practice trials containing all 10 phrases, each paired with a different dot position. The session lasted approximately 40 min.

Results and Discussion

Two pieces of information, a categorical response and a RT, were recorded on each trial. Subjects' RTs for each phrase-position pair (a maximum of three) were first grouped by response (true or false) and then averaged. If, for a particular phrase-position pair, a subject gave only one "true" or "false" response, the single RT was used. These means were then collapsed across subjects, keeping "true" and "false" responses separate. The resulting means were then graphed if more than half of the subjects contributed at least one RT to a mean. The procedure was followed to show RT effects for categorization in border regions where categories overlapped, but to exclude unreliable points based on only a few RTs. In all of the figures that follow, solid lines indicate those means based on responses from at least half of the subjects.

As in Experiment 1 there were some subjects for whom the categories "high" and "low" were bounded and did not extend to the extremes of the rectangle. However, in this experiment, there were only 2 such subjects out of the 20, and we did not include their data in the analysis. Because the small number of RTs these two subjects contributed did not yield sufficiently reliable means for a separate analysis, the responses of these subjects will not be discussed further.

The mean RTs for "high" are shown connected by the solid lines in the upper left panel of Figure 3. As expected, subjects took longest to respond "true" or "false" when the dot position fell near the border of the category (i.e., Position 5 for "high"). For "trues," when the dot position was near the category boundary, subjects took longer to respond. Similarly, for "falses," subjects took less time to reject category membership

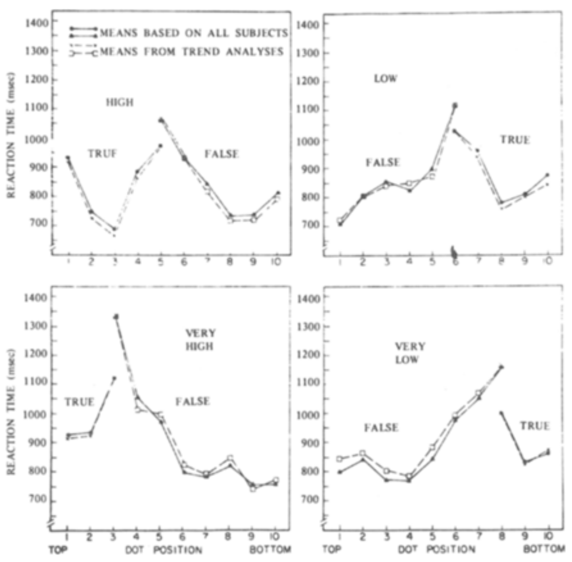

Figure 3. Mean reaction times from Experiment 2 for "high" (upper left panel), "very high" (lower left panel), "low" (upper right panel), and "very low" (lower right panel).

when the items were further away from the boundary toward the bottom of the rectangle.

There also were some unexpected effects in the "true" responses. There was the anticipated increase in mean latency as the dot position moved from Position 3 to Position 5, but there was also a marked increase in mean latency going from Position 3 to Position 1. Subjects took much longer to respond "true" to Position 1, the highest dot position, than to Position 3, even though all 18 subjects responded "true" to both positions. To test the statistical significance of the V-shaped function for the "high trues," a trend analysis was performed on the data from those 11 subjects who contributed (true) responses to all positions, 1 through 5. While this method of analysis does exclude some data for Positions 1-4, it allows a within-subjects analysis for trend components. For the statistical analyses that follow, an effort was made to include response latencies for items in the fringes of categories. The means based on just the data from the 11 subjects contributing "true" responses to all positions, 1 through 5, are shown by a dotted line in the upper left panel of Figure 3. Note that the means for the 11 subjects (connected by the dotted line) are very close to those for all of the subjects together (the solid line). The same convention will be used in all the graphs that follow: Dotted lines will show the means for data included in trend analyses, the results of which will be presented in tables. Table 1 contains a summary of the trend analyses from Experiment 2, among the most important of which was the reliable quadratic trend in the "true" means $[F(1,40) = 20.77, p < .001]$.

Results for "very high" are shown in the lower left panel of Figure 3. In general, subjects responded more quickly to dot positions that were further from the category boundary. The modifier "very" caused a restriction in extension relative to the base concept "high," but there was a minor departure from the results of the first experiment. While the membership function from Experiment 1 included Position 4 as a marginal member, a majority of subjects in this experiment did not consider it within the category "very high."

The upper right panel of Figure 3 shows the mean latencies for "low." Again, subjects responded faster as the dot position moved away from the category boundary toward the lower end of the rectangle. It is important to notice that the data from "true" responses to "low" contained a pattern similar to that for the "high trues." Subjects responded fastest to Position 8, the middle of the category, and slower to Position 10, even though subjects were in complete agreement. As suggested by the graph, the data from the 14 subjects giving responses for Positions 6-10 contained a reliable quadratic component $[F(1,52) = 7.73, p < .01]$. "False" responses for "low" were generally slower when the dot position was further from the top of the rectangle and closer to the category fringe.

The means for "very low," shown in the lower right panel of Figure 3, are very similar to those for "very high." The shift in the extension of the base term "low" caused by "very" was, in most respects, identical to the shift produced for "very high." It caused a restriction of two dot positions, and there were large effects of uncertainty. As expected, subjects took longer to decide category membership for items falling on the fringes of categories, and they gave inconsistent responses about some items' membership. These findings appear to be in agreement with the grade-of-membership functions obtained in the first experiment.

There are at least two explanations for the V-shaped functions obtained for RTs to "high" and "low." First, the RTs may have shown a large effect of category prototypes. If one uses Reed's (1972) definition of a category prototype as that item possessing the average values on all relevant stimulus dimensions, then in the present situation a category prototype would be defined by a single average on the height dimension. The

Table 1
Summary of Trend Analyses for Experiment 2

		Trend Component		
Phrase	Response	Linear $p <$	Quadratic $p <$	Residual $p <$
High	True	n.s.	.001	n.s.
	False	.001	.001	n.s.
Very High	True	.001	.10*	
	False	.001	.001	.025
Low	True	.01	.01	n.s.
	False	.001	.01	.01
Very Low	True	.025	.025	
	False	.001	n.s.	.01

*$.05 < p < .10$.

prototypical "high" would be computed by finding the average dot position of all those included in the extension of the category. Most subjects included the top five positions in "high," and, since each position occurred equally often in the experiment, the middle dot position (Position 3) would be the most representative member. Position 3, in fact, had the fastest mean RT for "high." Similarly, the prototype for "low," based on the five dot positions included by most subjects, would be Position 8, and subjects classified a dot position as a member fastest when it was at Position 8. When interpreted in this way, these results bear a striking resemblance to typicality effects reported in the semantic memory literature (Caramazza, Hersh, & Torgerson, 1976; Rips, Shoben, & Smith, 1973). Subjects in these studies could assign membership to typical or highly representative category members faster than they could to atypical members.

The interpretation presented above suggests that subjects used category prototypes to make their judgments, but the membership functions for the three linguistic subjects in Experiment 1 suggest a different account for the obtained V-shaped functions, one based on response competition. Recall that, for the linguistically responding subjects, the categories "high" and "low" were bounded on either side. A dot position that was best described by "very high" was not also "high," and a "very low" dot position was not "low." These subjects' membership function for "high" had a peak at Positions 3 and 4, but dropped off and did not include Position 1 as a member, and the corresponding function for "low" had a peak at Position 8 but again dropped off, and Position 10 was not part of the "low" extension. In general, the membership curves showed that an item was assigned membership in a category only if there was not a more appropriate or accurate descriptor for that position in the set of experimental phrases. In Experiments 1 and 2, "high" was bounded by "very high" on one side and "sort of high," one of the context phrases, on the other; "low" was bounded by "sort of low" and "very low." Similarly, "very high" and "very low" were bounded by "extremely high" and "extremely low," respectively. Referring to the RT graphs for "high trues" and "low trues," one can see that when a dot position can be described by a different phrase, the latency to respond increases. Mean RT for "high," for example, is fastest at Position 3, where "high" is the best descriptor. At Position 2, where "very high" is a better descriptor, subjects still respond "true" but they take longer to do so. And at Position 1, which is best described by "extremely high," subjects take still longer to respond. The same interpretation applies to the data for "low."

EXPERIMENT 3

The results from Experiment 2 allow, therefore, at least two interpretations. One interpretation is based on subjects' use of category prototypes, and the other holds that the results simply reflect the effect of response competition. To choose between these two interpretations of the results of Experiment 2, the speeded categorization paradigm was used again, but with the obvious sources of response competition removed. The phrases "very high" and "very low," as well as all six context phrases, were dropped from the stimulus set, leaving only the two base terms "high" and "low." Thus, the stimulus set contained two categories that were (almost) mutually exclusive. Except for the dot heights falling near the boundary between the two categories "high" and "low," a pictorial stimulus could not have simultaneous membership in more than one category. If category prototypes provide the basis for a general strategy subjects use to categorize pictorial stimuli, then similar effects should obtain in this simplified experimental context. Alternatively, if some form of response competition was at work in Experiment 2, then the V-shaped RT function should not obtain with the less complex stimulus set.

Method

Subjects. Twelve Johns Hopkins students who had not participated in the first two experiments served as subjects.

Stimuli. Slides of two phrases, "high" and "low," were paired with slides of 10 different dot positions. There were 9 replications of the set of 20 phrase-position pairs, and the resulting total of 180 trials was presented in a random order.

Procedure. The procedure was identical to that used in Experiment 2, except that subjects responded by pressing either a "true" button or a "false" button on a hand-held response box. The hand used to answer "true" was counterbalanced across subjects. Before starting, subjects saw 20 practice trials that included one example of each phrase-position pair in a random order.

Results and Discussion

The data were analyzed as described in Experiment 2. An individual subject's RTs for each phrase-position pair (a maximum of nine) were first grouped by response and then averaged. These means were collapsed across subjects, keeping "true" and "false" responses separate. As in Experiment 2, means were graphed if more than half the subjects contributed at least one RT to a mean. Of particular interest is the comparison between the present results and those for "high" and "low" in Experiment 2.

The left panel of Figure 4 shows the means for "true" responses to "high," and the trend analyses are summarized in Table 2. A majority of subjects responded "true" to "high" over Positions 1-5, just as in Experiment 2. However, unlike the RT function for "high trues" from the second experiment, these data contained no V-shaped pattern. The reliable quadratic component in these data instead most likely reflects a floor effect, since mean RT for Positions 1-3 are roughly equal. The "false" data for "high" are quite similar to those from the earlier speeded categorization task.

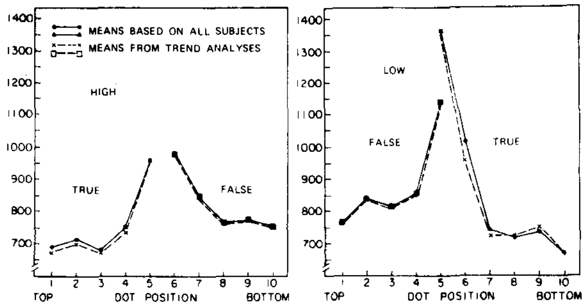

Figure 4. Mean reaction times from Experiment 3 for "high" (left panel) and "low" (right panel).

The right panel of Figure 4 shows the means for "low," and the relevant trend analyses are summarized in Table 2. The data for "low" from this experiment show the same pattern as those for "high." For both "true" and "false" responses, mean RTs were longer near the categories' boundary. At Position 5, more than half of the subjects responded "true" and "false," and for Dot Position 5, further away from the category boundary, subjects answered more quickly. Also, the "low true" responses did not contain any hint of a V-shaped pattern, such as that obtained in Experiment 2.

In summary, when subjects were confronted with the reduced stimulus set used in Experiment 3, their RTs still showed large effects of uncertainty, but there was no suggestion of a prototype effect such as that obtained in Experiment 2. Instead, the data supported the response competition hypothesis. When the only category labels available were "high" and "low," subjects responded equally quickly to all dot positions except those that fell on the fringe between the two categories.

EXPERIMENT 4

The argument presented thus far rests upon the comparison between Experiments 2 and 3; in one study, there were V-shaped RT patterns, and in the other, there were not. On the basis of this difference, a response competition hypothesis was proposed to account for the "prototype" effect. In order to make this point stronger, the apparent prototype effect predicted by the competition hypothesis should be replicated. For this purpose, four phrases, "high," "sort of high," "sort of low," and "low," were included in a set of category descriptors. Two of the phrases, "high" and "low," applied to the extremes of the rectangles, and the other two, "sort of high" and "sort of low," applied to the middle regions. The result of this ordering from top to bottom (high, sort of high, sort of low, low) was that the extensions for the phrases "high" and "low" overlapped with those from the adjacent categories. The overlapping extensions for the categories "high," "sort of high," "low," and "sort of low" provided a potential for response competition, as in Experiment 2. Specifically, it was hypothesized that subjects would take longer to respond "true" to "high" as the dot position shifted from the upper end of the rectangle toward the middle, where "sort of high" was a more appropriate descriptor, and that subjects would take longer to respond "true" to "low" when the dot position was more accurately described by "sort of low." Also, subjects should take longer to respond "true" to "sort of high" and to "sort of low" when other descriptors were more appropriate. Because the category "sort of high" is bounded on one side by "high" and on the other by "sort of low," mean "true" response time should increase on either side of some "prototypical" dot position. Similarly, "sort of low" is bounded on one side by "low" and on the other by "sort of high," and mean "true" RT to "sort of low" should increase when the dot position is better described by an adjacent category label.

Method

Subjects. Twenty Johns Hopkins students who had not participated in Experiments 1-3 served as subjects.

Stimuli. Slides of four phrases, "high," "sort of high," "low," and "sort of low," were paired with slides of 10 different dot positions. Five replications were made of each of the 40 phrase-position pairs, and the resulting 200 trials were presented in a random order.

Procedure. The procedure was the same as described under Experiment 3. Before starting the session, subjects saw 20 practice trials containing equal numbers of the four phrases and 10 dot positions in a random order.

Results and Discussion

The same method of analysis used in Experiments 2 and 3 was used here. Each subject's RTs for a phrase-position pair (a maximum of five) were first grouped by response (true or false) and then averaged. These means were collapsed across subjects, keeping "trues" and "falses" separate.

The upper left panel of Figure 5 shows the mean RTs for "high" and the lower left panel shows those for "sort of high." For both of these categories, there are effects of uncertainty similar to those found in the earlier experiments. Subjects did not always respond in a consistent manner to several phrase-position pairs on the fringes, and, as the dot position moved out of a category fringe, subjects made faster responses. For "high," subjects responded "true" fastest at Position 1 and progressively more slowly when the dot position

Table 2
Summary of Trend Analyses for Experiment 3

		Trend Component		
Phrase	Response	Linear $p <$	Quadratic $p <$	Residual $p <$
High	True	.001	.01	n.s.
	False	.001	.001	n.s.
Low	True	.001	.01	n.s.
	False	.001	.001	.001

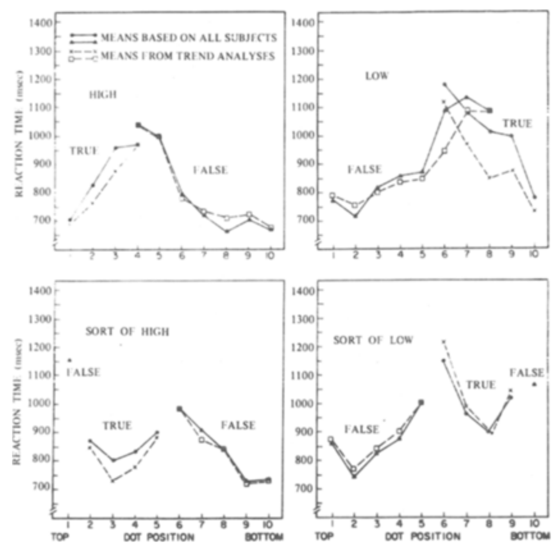

Figure 5. Mean reaction times from Experiment 4 for "high" (upper left panel), "sort of high" (lower left panel), "low" (upper right panel), and "sort of low" (lower right panel).

was closer to Position 4. They responded "false" progressively faster from Position 4 to Position 8 and continued to respond quickly to Positions 9 and 10. The presence of "sort of high" as an alternative category label in the stimulus set may cause a restriction in the category "high" from five positions to four; however, the importance of this restriction in and of itself is not clear, since the analogous restriction is not present in the "low" data. At Position 1 most subjects responded "false" to "sort of high," and subjects responded "true" fastest to "high" at Position 1. When the dot position was further down in the rectangle and "sort of high" was a more appropriate descriptor, subjects responded progressively more slowly to "high." This difference between the data for "high trues" from the present experiment and those from Experiment 2 is almost certainly not the result of a change in the meaning of the terms used, but simply the effect of response competition. Now, Dot Positions 2-4 can be described both as "high" and "sort of high." The function for "high falses" appears qualitatively very similar to that obtained in Experiment 2.

The "true" means for "sort of high" are shown in the lower left panel of Figure 5. The mean latencies based on the 17 subjects contributing responses to Positions 2-5 did not differ significantly from each other [$F(3,48) = 2.24$, $.05 < p < .10$]; however, they do invite comparison to the V-shaped function for "high trues" in Experiment 2. In the present experiment, the category "sort of high" was bounded on both sides: by "high" at the top of the rectangle and by "sort of low" toward the bottom. As alternative descriptors became more appropriate than "sort of high," the time subjects needed to respond "true" to "sort of high" increased. Support for this pattern is clearer in the data from "low." The mean "false" RTs followed the pattern shown in previous graphs.

The mean latencies for "low" and "sort of low" are shown in the right panels of Figure 5. Subjects gave "true" responses to "low" for Positions 6-10 and "false" responses to Positions 1-5. The presence of "sort of low" did not produce any restriction of the category "low" relative to Experiments 2 or 3. In other respects, the data for "low" and "sort of low" are equivalent to those for "high" and "sort of high." Subjects responded "true" more slowly to "low" when the dot position was further away from the bottom of the rectangle and close to the fringe of the category. Like the "high trues," the means for "low trues" did not show any V-shaped pattern. Subjects took longer to respond "true" to "low" when "sort of low" became a better descriptor, as the dot position shifted from the lower extreme of the rectangle to the middle. The "low falses" showed a regular increase in mean RT as the dot position moved from the top of the rectangle to the category fringe. There was, however, an unexplained dip at Position 8.

The "false" data for "sort of low" resemble those for "sort of high," in that there are steady increases in RT as the dot position moves closer to the fringe of the category. The "true" responses showed similar increases in mean response latency when the dot position fell close to the boundary of the category on either side. Also, as predicted by the response competition hypothesis, the "true" means contained a reliable quadratic component [$F(1,48) = 7.37$, $p < .01$]. Results from this and other trend analyses are shown in Table 3.

The important results for present purposes are the means for "true" responses to the four phrases "high," "sort of high," "low," and "sort of low." The functions for "high" and "low" support a response competition explanation for the prototype effects obtained in Experiments 2 and 4; as the dot position moved away from either extreme of the rectangle toward the center, subjects responded (true) gradually more slowly due to increasing interference from the deintensified category

Table 3
Summary of Trend Analyses for Experiment 4

Phrase	Response	Trend Component: Linear p <	Quadratic p <	Residual p <
High	True	.01	n.s.	n.s.
	False	.001	.01	n.s.
Sort of High	True	n.s.	n.s.	n.s.
	False	.001	n.s.	n.s.
Low	True	.001	n.s.	n.s.
	False	.001	.05	n.s.
Sort of Low	True	.10*	.01	n.s.
	False	.001	.001	n.s.

*$.05 < p < .10$.

labels "sort of high" and "sort of low." Similarly, the quadratic functions for "sort of high" and especially "sort of low" reflect increasing amounts of response competition from adjacent categories with partially overlapping extensions.

GENERAL DISCUSSION

The purpose of the three speeded verification experiments was to examine how subjects decided category membership for pictorial stimuli when those stimuli could belong to more than one category included in the experimental context. Notice that except for a minor restriction of the category "high" in Experiment 4, the "meanings" for the categories were the same in all studies and did not change as a function of the experimental context. In Experiment 2, there were a number of alternative category labels for dot positions falling within the extension of the category "high" and within that for "low," and the presence of these descriptors caused response competition. Subjects were able to verify membership more quickly for items that fell at the center of the target category specified on a trial (i.e., those items that were closest to subjects' best idea or image of a category's meaning).

In Experiment 3, in which the only category concepts included were "high" and "low," subjects were able to adopt a simpler approach to the task. On receiving a target category on a trial, a subject might have focused on the midline of the rectangle. The further away a dot was from the region of uncertainty, that is, the boundary of the categories, the easier it was to verify or reject category membership. The important point here is that subjects might logically have used this same strategy in Experiment 2, since the definitions for the two category concepts (high and low) were the same in the two studies. Similarly, in Experiment 4, subjects could again have used a simple rule for deciding category membership in "high" and "low." However, the mean response latencies were influenced by response competition from "sort of high" and "sort of low." The results from this study showed gradually decreasing functions for "true" responses for "high" and "low," rather than the floor effects obtained in Experiment 3.

Taken together, the results from these experiments provide an important demonstration of how subjects interpret and process dimensional adjectives. A necessary prerequisite for adequate interpretation of dimensional adjectives is a norm to which the poles of the scale can be related. Bierwisch (1967) points out that such a norm is often supplied by the semantic description of the object being described. For example, the sentence, "The man is sort of short," is more or less precisely interpretable given an understanding of the meaning of "man." Ambiguity is possible, of course, but if a particular group of men, deviating from the height norm for men in general, was intended as the reference group (e.g., professional basketball players), the burden would fall on the speaker to clarify that norm. One would expect the use and interpretation of these forms to conform to Grice's (1975) maxims of conversational implicature; that is, all of the information needed for interpretation must be supplied, but no more. It is intuitively evident that this maxim is generally followed by adults, and that successful (and efficient) communication about the size of various objects is easily accomplished by the dimensional adjectives and various modifiers, with further contextual information supplied only if it is necessary. The results of Experiments 1, 2, and 4 follow Grice's scheme. Thus, an interpretation of the response competition found in Experiment 1 is that the subject's first inclination is to respond "true" only to those phrase-picture pairs where the phrase is clearly the most appropriate descriptor for the picture. For example, when a subject is given the pair "very high," Position 2, he responds "true" quickly because the phrase "very high" is an accurate and, more importantly, pragmatically appropriate descriptor of Position 2. However, when the phrase "high" is paired with Position 2, it is still "logically" correct but pragmatically inappropriate because of the availability of the phrase "very high." It would appear, then, that a complete description of category membership extends beyond those aspects that are specified in an intensional definition of a concept to include pragmatic aspects of language use.

Implications for Related Work

One very important factor affecting the categorization process in the experiments described above concerns the nature of the category boundaries: Category boundaries are vague rather than precise. The property of vagueness can be illustrated with the following example. A man who is 4 ft 10 in. tall is clearly not a tall man, at least relative to American males today. If 1/16 in. is added to his height, he will still not be tall. If a second 1/16 in. is added, he will still be less than tall. Yet, if this incrementing is continued, the man will eventually become tall. Given this end result, the problem of specifying the meaning of the concept tall—the boundary for the category "tall men" in this context—reduces to identifying the single increment of 1/16 in. that will make the difference between membership and nonmembership in the category "tall." Unfortunately, identifying the single, criterial increment is not possible. Category boundaries are not precise; instead, there is a range of values on the height dimension, from roughly 5 ft 8 in. to 6 ft 0 in., for which there is not a clear and discrete distinction between membership and nonmembership. The fringe or region of uncertainty associated with a category reflects the vagueness of the category concept. Vagueness cannot be adequately explained as being a result of perceptual confusion alone; for example, even if one knew exactly how tall an individual was, that person's height could still fall in a region of vagueness.

The property of vagueness has attracted the attention of philosophers (e.g., Black, 1937; Russell, 1923), linguists (e.g., Labov, 1973; Lakoff, 1973), and psychologists (e.g., Hersh & Caramazza, 1976; Neisser, 1967). A seminal paper in the investigation of this problem is one by Labov (1973), who studied the categorization (naming) of container-like objects (cups, bowls, glasses, etc.) that varied on several dimensions either discretely (handles vs. no handles) or continuously (width and depth). As the depth of the objects increased relative to the width, subjects used the label "glass" more and more often and "cup" less often. As the width of the objects increased relative to the depth, more "bowl" responses were given relative to "cup" responses. The important point here is that some of the objects did not look like a good cup or a good bowl or a good glass; instead, they fell somewhere in between the different categories. When subjects described these ambiguous objects, they produced a mixture of category labels (e.g., some "bowl" and some "cup"). That is, there were physical stimuli that did not possess clear membership in any one category, but instead possessed partial membership in more than one category. Labov has interpreted the existence of regions of indeterminacy as support for the notion that natural language categories are vague.

Hersh and Caramazza (1976) have examined the property of vagueness using the dimensional adjectives "large" and "small," both alone and in combination with adverbial modifiers such as "very," "extremely," "sort of," and so on. These category descriptors label obviously vague concepts and therefore provide a convenient context for studying this property of natural language concepts. There were clear effects of vagueness for the dimensional adjective concepts tested similar to those reported by Labov (1973) for noun concepts. For certain phrase-square combinations, subjects did not respond in a consistent manner; and for individual subjects as well as for groups of subjects, there were fuzzy boundaries for the categories used in the experiments. In addition, the adverbial modifiers (very, sort of, etc.) did not affect the inherent vagueness of the base concepts. Adverbial modification did affect the extension of the concepts, however. "Very," for example, restricted the extensional definition of both "large" and "small," but such adverbial modification did not make category boundaries (e.g., those of "very large") any more or less vague or fuzzy than those for the base concepts (e.g., large).

The present work replicates the results of Hersh and Caramazza (1976) with respect to the effects of adverbial modification on dimensional adjective concepts. Also, the experiments reported here examined the real-time processing effects of vagueness on categorization. When a pictorial stimulus was in the fringe area of a category, subjects had difficulty (i.e., produced longer RTs) deciding the object's membership. This result of uncertainty at category boundaries can also be explained as a result of perceptual confusion. Such an account would be as follows. Category boundaries might be precise and well defined, but subjects have trouble judging exactly where the pictorial stimulus falls. Thus, the increase in RT associated with dot heights near category borders is simply a result of perceptual confusion. Alternatively, the RT effects of uncertainty are due primarily to the vagueness of the underlying representation of the category concepts rather than to perceptual confusion (Hersh & Caramazza, 1976). Unfortunately, however, it is difficult to distinguish between the vagueness explanation and the perceptual confusion account on empirical grounds.

REFERENCES

Bierwisch, M. Some semantic universals of German adjectivals. *Foundations of Language*, 1967, **3**, 1-36.

Black, M. Vagueness. *Philosophy of Science*, 1937, **4**, 427-455.

Caramazza, A., Hersh, H. M., & Torgerson, W. S. Subjective structures and operations in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 103-117.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240-247.

Grice, H. P. Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts*. New York: Seminar Press, 1975.

Hersh, H. M., & Caramazza, A. A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General*, 1976, **105**, 254-276.

Labov, W. The boundaries of words and their meanings. In C. J. N. Bailey & R. W. Shuy (Eds.), *New ways of analyzing variation in English*. Washington, D.C: Georgetown University Press, 1973.

Lakoff, G. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 1973, **2**, 458-508.

Neisser, U. *Cognitive psychology*. Englewood Cliffs, N.J: Prentice-Hall, 1967.

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

Reed, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, **3**, 382-407.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 1-20.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, **8**, 382-439.

Russell, B. Vagueness. *Australasian Journal of Psychology and Philosophy*, 1923, **1**, 84-92.

Smith, E. E. Theories of semantic memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 6). Potomac, Md: Erlbaum, in press.

Zadeh, L. A. Fuzzy sets. *Information and control*, 1965, **8**, 338-353.

(Accepted for publication August 24, 1978.)