JOURNAL OF EXPERIMENTAL SOCIAL PSYCHOLOGY **33**, 122–145 (1997)
ARTICLE NO. JS961315

Effects of Arbitrarily Placed Category Boundaries on Similarity Judgments

MYRON ROTHBART AND CARENE DAVIS-STITT

University of Oregon

AND

JONATHAN HILL

University of Minnesota

Received: August 28, 1995; revised: May 16, 1996; accepted: July 17, 1996

Three experiments examined the effects of arbitrarily placed boundaries on judged similarity between pairs of stimulus persons. Subjects in Experiments 1 and 2 rated the similarity of applicants for a job in which three category labels (Ideal, Acceptable, and Marginal) were superimposed on the composite scores of applicants based on measures relevant to job performance. The exact position of the category boundaries, described to the subjects as arbitrary, was varied across two conditions, allowing a comparison of a given pair of stimulus persons when the members of the pair were in the same category, or in different categories. Whereas the boundary positions were varied across subjects in Experiment 1, they were varied within subjects in Experiment 2. Subjects in both experiments rated pairs as more similar when in the same category than when on opposite sides of a category boundary. In Experiment 3, subjects rated pairs of movie actors on three scales assessing dispositional similarity, based upon the degree of the actor's political liberalism on a percentile scale. The presence of semantic labels, as well as visual boundary markers, was varied, and again confirmed the predicted effects of boundary markers. The presence of a no-categorization control indicated the presence of both intercategory accentuation and intracategory minimization. Moreover, actors separated by objectively large distances *within* a category were seen as more similar than actors objectively smaller distances apart but separated by a category boundary. © 1997 Academic Press

This research was supported in part by National Institutes of Mental Health Grant MH40662 to the first author. We gratefully acknowledge the help of Deborah Day in running the experiments, the editorial comments of Mary Rothbart, and additional comments by Jochen Krueger and Chick Judd. Reprints may be requested from Myron Rothbart, Ph.D., Department of Psychology, 1227 University of Oregon, Eugene, OR 97403-1227.

0022-1031/97 $25.00

There is an old Yiddish story of a peasant whose farm was located near the border of Poland and Russia, where boundary markers shifted with every international dispute. The peasant did not know from one year to the next whether his farm was in Russia or Poland, and eventually hired a surveyor to resolve the uncertainty. After weeks of painstaking assessment, the surveyor finally announced that the farm was just inside the Polish border. "Thank God," the peasant cried with relief, "now I won't have to endure any more Russian winters!"

The irony in the story resides, of course, in the peasant's belief that an arbitrarily placed national boundary, shifting unpredictably with political fortunes, was a more important indicator of climate than the fixed, physical location of the farm. The underlying "reality" in this case is the prevailing climate at the farm's location, not the arbitrary national boundaries superimposed on the region. Although the story could be dismissed as yet another example of human folly, it may provide a useful glimpse into the more general character of category labels.

The power of category labels has been long recognized by social psychologists. Allport, in *The Nature of Prejudice,* noted the importance of labeling in the process of social perception, and the ability of some labels to inhibit alternate bases for classification (Allport, 1954). Rosenhan, in his classic demonstration of the power of the label "mentally ill," illustrated the potential "sticking power" of a label that was difficult to disconfirm, particularly in the context of institutional norms and behavior (Rosenhan, 1973). Nevertheless, the process of labeling is not a unitary phenomenon.

The power that Allport and others ascribe to labels may be due in large part to the strong meaning associated with particular labels (for example, Communist, handicapped, schizophrenic, Republican), rather than to the phenomenon of "being labeled." This is a subtle distinction, and deserves elaboration. Labels often provide useful information, but frequently they do not. Consider the example of the psychoanalyst who, in deference to health insurers' needs for diagnostic categories, routinely describes his clients as "Psychoneurotic, Mixed Type." Since few would fail to qualify for membership in this category, it comes close to being a pure case in which the label itself adds little substantive information to the person being labeled. Distinctions between psychiatric categories are often assumed to reflect meaningful intercategory discontinuities, but the placement of boundaries for other types of categories are thought to be arbitrary. The classic example of color naming, in which regions of the visible portion of the electromagnetic spectrum are given color names, is an example of an ordinal classification where the exact placement of boundaries is assumed to be arbitrary (Brown, 1976). Although the category boundaries that are used to separate regions along a continuum are often recognized as arbitrary, they nonetheless may influence the perceived similarity of objects along that continuum.

In reality, it is difficult to separate the meaning associated with a label from the effects of labeling itself, since the function of a label is to add meaning to the objects subsumed under the category name. A classic set of studies by Tajfel and Wilkes (1963) examined the effects of labeling on the perception of ordered

objects using an original paradigm. A series of eight lines, graded to increase by a fixed ratio, were presented to subjects in a number of conditions. In a random labeling condition, the letters A or B were randomly associated with each of the eight lines, and in another condition, each letter was associated with either the four shortest or the four longest lines. In the latter condition the letters, in effect, took on the meaning of "short" or "long." Two predictions were made regarding subjects' estimates of line length: (*a*) the difference in estimates between the two lines at the category boundaries (lines 4 and 5) should be greater under the meaningful than the random labeling condition and (*b*) differences in estimates within categories (1–4 and 5–8) should be less in the meaningful than random labeling conditions. In other words, they predicted intercategory accentuation (contrast), and intracategory minimization (assimilation).

Although the authors claimed support for both hypotheses, there only seemed to be support for intercategory accentuation in their data, and no support for within-category assimilation. The Tajfel and Wilkes findings show evidence of intercategory accentuation, but produced by a constant displacement of the values of one category away from the values of the other, rather than by the "shrinking" of values within each category toward its own central tendency.[1] More generally, intercategory accentuation can in principle be produced by at least two distinct processes: (*a*) the boundary demarcation itself leads to a displacement of one distribution away from the other, or (*b*) within-category estimates are assimilated to a central tendency, resulting in all values being drawn to a central value, and incidentally to an accentuation of perceived differences at the boundaries of the category.

In a very clever experiment, Krueger and Clement (1994) were able to test these two possibilities in a domain familiar to our Polish peasant. They asked undergraduates to estimate the average local temperature of specific days of the year in which the dates, separated by 8-day intervals, were presented in random order. They were thereby able to examine the perceived differences between adjacent 8-day periods both within and across monthly boundaries. Differences between perceived temperatures across these fixed intervals were seen as greater when the 2 adjacent days appeared in different months compared to 2 adjacent days appearing in the same month. Krueger and Clement were further able to test the two alternative interpretations of category accentuation. Using regression techniques, it was possible to determine whether the predicted temperatures at days near the monthly boundaries could be adequately predicted from the regression lines within each month, or whether the slopes under- or overestimated the predicted temperatures. Their findings strongly supported an assimilation model in which temperatures were assimilated to monthly averages, with no

[1] The beauty of the Tajfel and Wilkes design is that it allowed a comparison of categorized values with "reality" (i.e., uncategorized values). An inspection of their findings shows that the distribution of larger values is displaced upward, but there is no downward displacement of the smaller values. The constant displacement model does not necessarily assume that both distributions will be displaced, but of course assumes that at least one is displaced.

evidence that the monthly boundaries themselves were contributing to an accentuation effect.

The studies by Krueger and Clement asked subjects to provide estimates from memory, and there is little question that subjects were retrieving their accumulated knowledge of seasonal variations in temperature to make their estimates. Generally, subjects expected temperatures to increase from January to July, and to decrease from August to January, but more specifically they had clear expectancies that, for example, May is warmer than April, and October is cooler than September. When, for example, the temperatures for specific days in April are assimilated to the perceived central tendency for April, and the temperatures for specific days in May are assimilated to the perceived central tendency for May, the estimated temperatures for early April and early May are overestimated (by being judged closer to the average of the month than they ought to be), and for late April and late May are underestimated (by being judged closer to the average of the month than they ought to be). The discontinuities at the monthly boundaries are not due to the boundaries themselves, but due (in spring) to the underestimation of the last days of the earlier month and the overestimation of the first days of the later month. Thus, both the reduced variance within the category and the accentuation at the boundaries of the category are produced by assimilation to expected central tendencies.

Although the write-up is not as clear for the Tajfel and Wilkes research, it appears that their subjects' judgments were based on memory representations as well, where subjects retrieved, in the meaningful label condition, estimates based on their memories of the central tendencies for the "short" and "long" lines.

The analysis by Krueger and Clement suggests, ironically, that category accentuation effects in that study had little or nothing to do with category boundaries. Rather, category accentuation was a byproduct of assimilation to central tendency in which subjects significantly underestimated within-month variability around the average temperature. Although monthly boundary markers are arbitrary demarcations (Roman, Gregorian, Hebrew, Islamic, Aztec, and Mayan calendars use different demarcations), subjects were not using those arbitrary markers in making their judgments.

The goal of the present research is to examine the effects of arbitrarily placed boundaries on the perceived similarity of categorized objects. Whereas the Krueger and Clement study asked subjects to judge a naturally existing ordered set of objects, drawing upon real-world knowledge, the present set of studies experimentally manipulates the placement of boundaries, emphasizes their arbitrariness, and uses stimuli about which subjects have little or no prior information. The three experiments presented assess whether arbitrarily placed boundaries affect the perceived similarity of a given pair of objects when the members of the pair are or are not separated by an arbitrarily placed boundary. The difficulty, of course, is to manipulate the presence of boundaries without adding meaning to the interpretation of the objects being judged.

In all of the experiments to be reported, subjects are provided with continuous information on a dimension which, in itself, is the only "real" or meaningful information about the stimuli. Arbitrary boundaries are then placed along the continuum, and in Experiments 1 and 2, the arbitrariness of the boundary positions are explicitly described to subjects. Experiment 1 varies the placement of the boundaries across subjects and examines between-subject differences in the judged similarity of pairs of individuals when those stimulus persons are on the same side or on the opposite sides of an arbitrary boundary. Experiment 2 examines the same question, but with the arbitrary boundaries varied within-subject, rather than across subjects. Experiment 3, using a different set of stimuli, examines the joint effects of verbal labels and visual boundary markers on similarity judgments of pairs of movie stars whose "real" distance is systematically varied. Experiment 3 allows us to pit the effects of category boundaries against the "real" distance between stimulus persons.

STUDY 1

Overview

Subjects were given the task of rating the similarity of pairs of applicants for a managerial job. Composite scores, reflecting performance on job-related skills, were presented for seven female applicants, in which three category labels (Ideal, Acceptable, and Marginal) were superimposed on the continuous composite measure. The position of the category labels was described as arbitrary, varying unpredictably from year to year, depending on both the number of job applicants and the number of positions available. The exact position of the category boundaries was varied across subjects, allowing for comparisons of pairs of stimulus persons within and across category boundaries.

Method

Subjects. Seventy-six subjects, 30 men and 46 women, participated in this experiment in partial fulfillment of a research requirement for an Introductory Psychology course.

Stimulus materials. Each subject received a packet of materials containing eight file folders. Seven of the folders were the application files of the seven female candidates for the job (with the first name of the applicant listed on the tab), and the eighth file contained both a "summary sheet" for the applicants and the dependent measures. Each application file contained three pages of information. Stapled to the left page of the open file was a graphic depiction of the categories associated with the composite scores, a copy of which is shown in Fig. 1. The scale ranged from 500 at the bottom to 1000 at the top, and was associated with four labeled categories. Scores below 500 were marked with a descending arrow and the label Reject. In ascending order, the other three categories were Marginal, Acceptable, and Ideal. For Condition 1 the boundary between the Ideal and the Acceptable category was placed at 880, and between the Acceptable and the Marginal category was placed at 680. For Condition 2 the respective boundaries were placed at 820 and 620.

On the right side of the applicant's file were two pages, one over the other, stapled at the top. The top page contained the applicant's composite score in a box in bold letters in the center of the page. The page underneath contained a completed employment application form, listing the applicant's education, work experience, aptitude test scores, specific skills, and hobbies.

For each application file, the subject was to examine the file, enter both the numerical composite score and the appropriate category designation on the "summary sheet" using both the composite score and the graphic scale on the inside of each file. In addition, subjects were asked to write a

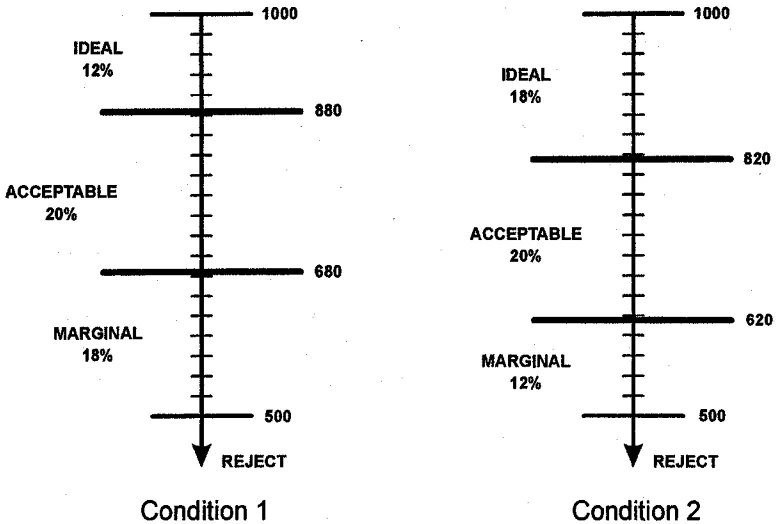

FIG. 1. Relation of composite scores to category labels, as presented to subjects.

one-sentence summary description of each applicant in terms of their appropriateness for the job. The applicant files were presented in random order, but the listing of names on the summary sheet occurred in a specific order, with the name of the applicant with the highest composite score at the top, and the remaining names in descending order based on their composite score.

After all seven files were examined, with the composite scores, category designations, and one-sentence description provided on the summary sheet, subjects rated the "similarity of pairs of individuals in terms of job performance" on a scale of 1 to 9 ("not at all similar" to "extremely similar") for 14 pairs of applicants.[2]

Table 1 presents the seven applicants, along with their composite score, and their category label as it varies by condition. Two of the applicants, 3 and 5, change categories as a function of different boundary positions. On the basis of these two changes, it was possible to generate six pairs of applicants who appeared as a within-category pair in one condition and a between-category pair in another condition: (1) 1 and 3, (2) 2 and 3, (3) 3 and 4, (4) 4 and 5, (5) 5 and 6, and (6) 5 and 7.

Procedure. The experiment was described to subjects as involving "the personnel selection process." Subjects were asked to read information about seven different "job applicants" and to note which category (Ideal, Acceptable, or Marginal) each person should be placed in. This decision was to be based on the person's composite score (which reflected the applicant's education, experience, aptitude test scores, and a personal interview) and a decision rule for determining the category boundaries for each condition. To make the arbitrariness of the boundaries clear, subjects were shown a table indicating the category boundaries in previous years and a large poster of a sample scale with sliding boundaries. Subjects were given the following tape-recorded information:

> The boundaries of the categories depends on two factors: the number of job applicants, and the number of jobs available at any particular time. The number of individuals placed in each category, therefore, is strictly arbitrary and varies from one time to the next. For example, in 1980 jobs were plentiful and job applications were down, so all of the job applicants whose composite scores were over 500 could be assigned to the ideal category. However, in 1981 there were many job applications and only a few jobs, so only applicants scoring over 900 were assigned to the ideal category.

[2] There were 21 possible pairs of applicants to be rated, but only 14 pairs were selected for rating. Only pairs close together were selected and pairs far apart were not rated. Of the 14 pairs, 6 pairs were critical; that is, the members of the pair were in the same category in one condition but in different categories in the other condition. The remaining 8 pairs served as filler items.

TABLE 1
COMPOSITE SCORE FOR SEVEN APPLICANTS AND CATEGORY LABEL BY CONDITION (EXPERIMENT 1)

		Category[a]	
Stimulus person name	Composite score	Condition 1	Condition 2
1. Joan	990	Ideal	Ideal
2. Claudia	915	Ideal	Ideal
3. Deborah[b]	840	Acceptable	Ideal
4. Shirley	750	Acceptable	Acceptable
5. Susan[b]	660	Marginal	Acceptable
6. Patricia	585	Marginal	Marginal
7. Ellen	510	Marginal	Marginal

[a] For Condition 1, Ideal = 1000–881, Acceptable = 880–681, and Marginal = 680–500. For Condition 2, Ideal = 1000–820, Acceptable = 819–620, and Marginal = 619–500.
[b] Applicants in different categories for two conditions.

Results

Each subject made similarity judgments for six pairs of applicants. For any given subject, three of the pairs involved members who were in the same category, and three pairs involved members who were in different categories. Since the boundaries differed for the two experimental conditions, the pairs within-category for Condition 1 were between-category for Condition 2, and the pairs between-category for Condition 1 were within-category for Condition 2. The results were analyzed in a 2 × 2 ANOVA, with Category Boundary (within vs across) as a within-subjects factor, and Condition as a between-subjects factor. Note that for any given subject, the within-category pairs are different from the between-category pairs, but by aggregating across subjects, each pair appears equally often as both within and between categories. The ANOVA revealed a significant Category Boundary effect [$F(1, 74) = 6.18, p < .02$], with the within-category pairs rated more similar ($M = 5.77$) than the between-category pairs ($M = 5.37$). There was a Category Boundary × Condition interaction [$F(1, 74) = 38.70, p < .001$], indicating that the two sets of items differed in degree of similarity. The effect of category boundary was greater for one set of items (Mean difference = .58) than for the other set of items (Mean difference = .22). There were no other significant effects.

Discussion

As hypothesized, subjects rated pairs more similar when they were within the same category (had the same label) than when the same individuals were on opposite sides of the category boundaries (had different labels). It should be noted that category membership had an effect even when (*a*) subjects were required to enter both the numerical composite score and the assigned category and (*b*) the arbitrary character of the category boundaries had been clearly explained both verbally and in tabular form. In effect, we have shown that arbitrarily defined category boundaries, when superimposed on a continuum reflecting the nonarbitrary characteristics of the different individuals, nonetheless affect the perceived similarity between those individuals. It is also true, however, that the arbitrariness of the category boundaries was not demonstrated in any dynamic fashion. For any given subject, each stimulus person was assigned to one and only one category. A more powerful demonstration of the effects of arbitrary category boundaries would be to show that for any given subject, the perceived similarity between a given pair of individuals changes as a function of shifts in category boundaries. This requires the category boundaries to be moved within subjects, rather than across subjects, and describes the basic design of the next experiment.

STUDY 2

Overview

Study 2 was a conceptual replication of Study 1, with the exception that the placement of category boundaries was manipulated within subjects rather than across subjects. Since it did not seem plausible to have subjects judge exactly the same stimulus persons under two different conditions—and still obtain differences in the ratings—the stimulus persons judged the second time were almost identical except for name, name of former employer, and absolute position on the composite scale. The composite scores used in Experiment 1 were shifted downward by 1–3 points for the first set of ratings (Condition 1) and upward by 1–3 points for the second set of ratings (Condition 2) (out of a maximum of 1000), with the numerical differences between stimulus persons remaining virtually the same for the first and second ratings.[3]

Method

Subjects. Ninety-seven subjects, 31 men and 66 women, participated in this experiment, satisfying in part a research requirement for a lower division introductory psychology course. Three subjects were dropped due to missing data.

Materials. Subjects received two packets of materials, each containing a set of seven file folders and corresponding answer sheets. The packets of materials used for Condition 1 were identical to those used in Experiment 1, except that the composite scores ranged from 988 to 507 (instead of 990 to 510).

[3] To reduce subjects' suspicion, we did not shift each pair by a constant amount. For four pairs (2 and 3, 3 and 4, 4 and 5, and 5 and 6), the absolute differences were identical in both conditions (75, 90, 90, and 75, respectively). For two of the pairs (1 and 3, 5 and 7), the absolute difference was not identical, but occurred in a direction counter to prediction. That is, pair 1 and 3 had a difference of 149 points in Condition 1 (where the applicants were in different categories) and 151 points in Condition 2 (where the applicants were in the same categories). Similarly, pair 5 and 7 had a difference of 152 points in Condition 1 (where numbers were in the same categories) and a difference of 148 points in Condition 2 (where the numbers were in different categories).

TABLE 2
COMPOSITE SCORE AND CATEGORY LABEL FOR SEVEN APPLICANTS BY CONDITION

Stimulus person	Name		Composite score[a]		Category	
	Condition 1	Condition 2	Condition 1	Condition 2	Condition 1	Condition 2
1	Joan	Karen	988	992	Ideal	Ideal
2	Claudia	Paula	914	916	Ideal	Ideal
3[b]	Deborah	Barbara	839	841	Acceptable	Ideal
4	Shirley	Marie	749	751	Acceptable	Acceptable
5[b]	Susan	Carolyn	659	661	Marginal	Acceptable
6	Patricia	Marjorie	584	586	Marginal	Marginal
7	Ellen	Jennifer	507	513	Marginal	Marginal

[a] For Condition 1, Ideal = 1000–881, Acceptable = 880–681, Marginal = 680–5100. For Condition 2, Ideal = 1000–820, Acceptable = 819–620, and Marginal = 619–500.
[b] Applicants in different categories for two conditions.

For Condition 2, the composite scores ranged from 992 to 513 and the names of the applicants and their previous employers were changed.

For Condition 1 and Condition 2 the same basic materials, procedures, and measures were used as in Experiment 1. Table 2 contains a summary of the names, composite scores, and categories used for the two sets of ratings.

Procedure. The experiment was described to subjects as involving "the personnel selection process." For Condition 1, the instructions were identical to those of Experiment 1. After completion of the tasks for the first set of materials, subjects were given a second, different set of materials and asked to repeat the process for a "new set" of seven job applicants applying for a different position within the same large corporation. Half of the subjects completed materials designated Condition 1 first, and half completed the materials designated Condition 2 first.

Results

The within-subjects design allowed us to do a 2 × 6 within-subjects ANOVA, assessing the effects of both category boundaries (within vs across category boundaries) and item pair (six stimulus pairs). That is, for every subject, a given pair appeared within a category in one set and across a category boundary in another set, allowing a completely crossed design (Category Boundary × Stimulus Pair) with each subject. There was a significant main effect of category boundaries [$F(1, 93) = 5.61$, $p < .02$], with higher similarity ratings given to a pair in the same ($M = 5.39$) rather than in different ($M = 5.13$) categories. Although there was also a highly significant effect of different pairs [$F(5, 465) = 15.70$, $p < .001$], with some pairs judged more similar than others, there was no significant interaction between category boundary and item [$F(5, 465) = 1.67$, $p < .2$]. There is no evidence that the effect of the boundary was significantly greater for some pairs than others. The within-subjects effects in Experiment 2 thus replicate the between-subject effects observed in experiment 1.

Discussion

One limitation of the first experiment concerned the effectiveness with which the arbitrariness of the boundaries was conveyed to subjects. Although a clear rationale was given in Experiment 1 for the arbitrariness of boundary placement, the manipulation of boundary position itself occurred between subjects, not within subjects. A more convincing way to show the arbitrariness of boundaries would be to present the same pairs (or virtually equivalent pairs) to subjects on two occasions, with a given pair appearing within the same category in one rating, but in different categories in another condition. This is the design employed in the second experiment, and again we found evidence of greater perceived similarity when the two stimulus persons were within the same category.

Unlike the Krueger and Clement research, and to some degree the Tajfel and Wilkes research, in this study subjects did not base their judgments on previously learned information which had to be retrieved from memory. Estimates of temperature in Krueger and Clement were based first on the general expectation of seasonal warming and cooling periods, and second on specific expectations associated with each month. Indeed Krueger and Clement found that category accentuation effects were due to assimilation to monthly expectations rather than to any effects occurring at the category boundary itself. In the Tajfel and Wilkes research, the same interpretation is possible, since subjects may have formed an estimate of central tendency based on prior exposure to the A lines and on the B lines before direct exposure to the judged stimulus. In the present experiments, subjects based their judgments on information available to them on a summary sheet, and were required to enter correctly *both* the numerical composite score and the appropriate category. These judgments were thus not based on information previously stored in memory, and it seems unlikely that subjects formed a prototype of what an Ideal, Acceptable, or Marginal applicant was like. We interpret our experiments as showing that category accentuation effects can be obtained by processes other than assimilation to central tendency.

The present data do not enable us to distinguish between inter- and intracategory effects. That is, the findings of Experiments 1 and 2 may represent (*a*) intercategory accentuation alone, (*b*) intracategory minimization alone, or (*c*) both inter- and intracategory effects. The lack of a no-categorization control condition does not allow us to isolate the locus of the accentuation effects. A third experiment was designed to clarify, and expand upon, the results of the first two experiments. Using a different dimension with known stimulus persons, and a no-categorization control condition, it allows us to look at the independent effects of two types of category boundary markers, and allows us to "pit" categorization processes against the reality of the underlying continuum. The no-categorization control allows us to test for both inter- and intracategory effects. Moreover, similarity along three broad, interrelated dimensions are assessed to move beyond a single index of similarity related only to the dimension being manipulated.

In the third experiment, subjects were asked to make similarity judgments of movie actors, where the stimulus persons are located at given distances along a continuum (political conservatism–liberalism). Verbal labels, as well as vertical lines, served as independent manipulations of the boundary markers, and pairs separated by a given distance could be compared under conditions of no

boundaries, labeled boundaries, visual boundaries, or both. This experiment allowed us to assess the power of category boundaries by comparing the perceived similarity of two persons a relatively small distance apart, but separated by a boundary, with persons a larger distance apart but existing within the same category. Finally, an ancillary hypothesis was tested that the subject's own position along the continuum should influence similarity judgments, predicting greater perceived similarity as the distance between the judge and the judged objects increases (Hovland & Sherif, 1961).

STUDY 3

Overview

Subjects made similarity judgments between 15 pairs of well-known male actors located at different positions on a percentile scale of political liberalism. Seven of the pairs varied by 10 percentile points on the scale, 4 varied by 15 points, and 4 varied by 35 points. The presence or absence of verbal labels ("conservative," "moderate conservative," "moderate liberal," and "liberal"), and the presence or absence of visual boundaries (heavy vertical tick marks) along the continuum, was varied across four independent groups of subjects. The design allowed us to (*a*) assess the effects of category boundaries on inter- and intracategory similarity, (*b*) pit the "reality" of the continuum against the implicit meaning of the category, and (*c*) assess the effect of the judge's own position. First, within the 10 and within the 35 point pairs, it was predicted that perceived similarity would decrease as a label and a boundary was interposed between the pairs. Second, although in the absence of boundaries, 10 point pairs would be judged as more similar than 15 point pairs, the effects of real differences would be minimized as the 10-point pairs became separated by labels and boundaries. Third, the judge's own position on the continuum should affect perceived similarity, with similarity increasing as the distance between the judge and the target pair on the continuum increased.

Method

Subjects

One hundred twelve subjects, 35 men and 77 women, participated in this experiment, in partial fulfillment of a research requirement for a lower division introductory psychology course. Subjects were randomly assigned to one of four conditions: a no boundary/no label condition, a no boundary/label condition, a boundary/no label condition, and a boundary/label condition.

Stimulus Materials

Stimulus persons. In contrast to the contrived stimulus persons used in the previous two experiments, we wished to use stimulus persons already familiar to subjects. It was necessary to choose stimuli who were somewhat familiar, but for whom there was at least some ambiguity about their degree of similarity. We chose well-known male movie actors as the stimuli. They were presented to subjects on a percentile scale representing the stimulus person's degree of conservatism–liberalism (e.g., "Clint Eastwood is more liberal than 31% of the population," "Alan Alda is more liberal than 86% of the population"). It was assumed that subjects may have some knowledge of the rough position of some actors on this dimension, but that it would be possible to present actors "exact" position on this scale without engendering surprise or suspicion. A percentile scale was used because it had at least interval properties, and would be clear in its interpretation to all subjects. One additional reason for choosing the liberalism dimension is that subjects could identify their own position on this dimension, allowing a test of the third prediction regarding the relation of the subject's own position to judged similarity.

Thirty pretest subjects judged the degree of liberalism or conservatism (using the percentile scale) of 31 well-known male movie actors. One group of 15 subjects rated the actors on conservatism and the other group of 15 rated them on liberalism. The ratings were averaged within each group, and the correlation was computed between mean conservatism ratings and mean liberalism ratings across the 31 actors. The correlation was highly significant ($r = -.62, p < .01$) indicating modest consensus on the conservatism–liberalism of the actors. Twenty-four actors were chosen from the array of 31 for use in this experiment. Actors were chosen from each region of the scale, and two actors were "assigned" to each of 12 scale positions.[4] From these 12 scale positions, the 15 pairs were constructed. Of the pairs 10 percentile points apart, 3 were placed so that boundaries/labels would intercede between the pairs, while 4 would always be in the same category. All of the four pairs 15 percentile points apart would always be in the same category. Of the pairs 35 percentile points apart, 2 pairs would have one boundary/label interposed, and 2 pairs would have two boundaries/labels interposed. For any given subject, the particular actor selected for each of the 12 scale positions was chosen randomly from the two actors assigned to each position.

Percentile scale. On the basis of a pretest experiment, in which half of the subjects made similarity judgments based on a percentile scale of conservatism (a high score indicated the actor was more conservative than *X*% of the population) and half made judgments based on a percentile scale of liberalism (a high score indicated the actor was more liberal than *X*% of the population), we found identical effects for both the conservative and the liberal forms of the scale. For simplicity, then, the present experiment used only a single scale of liberalism. All subjects were given the same information about how to interpret the percentile positions of the actors. For subjects in the no boundary, no label conditions, they were presented only with a continuum defined by 0 at the left end and 100 at the right end of the scale, with heavy vertical tick marks only at the ends of the scale. For subjects in the boundaries conditions, three more heavy tick marks were added at the 25, 50, and 75 positions.[5] For subjects in the label condition, four labels were placed under the four quartiles of the continuum: Conservative, Moderate Conservative, Moderate Liberal, and Liberal. The four experimental conditions are illustrated in Fig. 2.[6]

Procedure

Subjects were randomly assigned to one of the four experimental conditions, and individually tested in sound-proof cubicles containing a personal computer. A computer program, created in MEL (Micro Experimental Laboratory), presented each of the 15 pairs in random order. Each pair contained two actors, with the actor's name above an arrow pointing to a position on the continuum (see Fig. 2 for illustration). Within a given experimental condition, the same continuum with or without labels/

[4] The percentile scale positions were 4, 9, 19, 29, 34, 44, 54, 64, 69, 79, 89, and 94. The pairs separated by 10 points were 9 and 19, 19 and 29, 34 and 44, 44 and 54, 54 and 64, 69 and 79, and 79 and 89. The pairs separated by 15 points were 4 and 19, 29 and 44, 54 and 69, and 79 and 94. The pairs separated by 35 points were 9 and 44, 19 and 54, 44 and 79, and 54 and 89.

[5] For the boundary condition, a decision was made to include secondary tick marks at the decile positions to emphasize that these were normal scaling positions, thereby avoiding conferring any special status on the heavy tick marks. Note that if anything the inclusion of the secondary tick marks weakens the effects of the primary tick marks, because a pair might be within the category bounded by the heavy tick marks, but still separated by the secondary, lighter tick marks.

[6] Figure 2 illustrates three actors, with one pair staying within a category, and the other pair separated by a category boundary. In the actual experiment, only two actor's names appeared on a given trial.

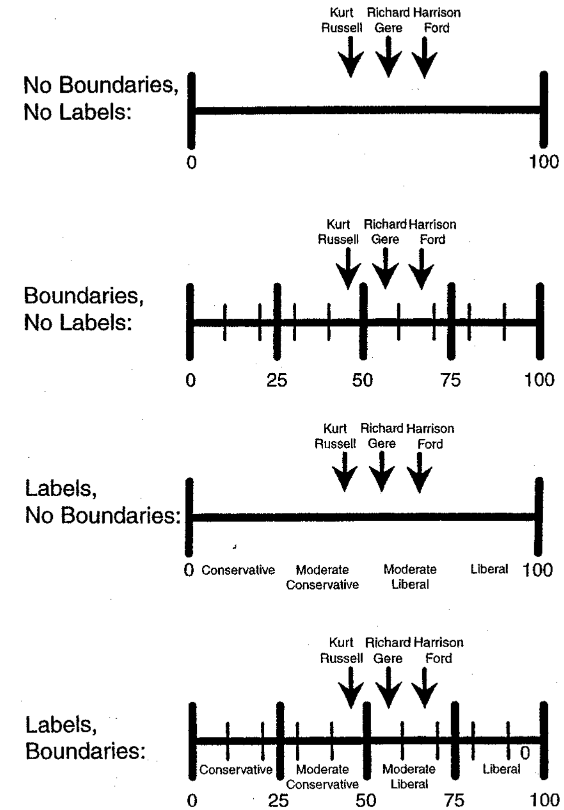

FIG. 2. Examples of stimulus presentation of pairs of actors by condition.

boundaries was used for all 15 pairs. The presence or absence of labels and boundaries varied across subjects, not within subjects. For each pair, the subject was asked to judge the degree of similarity between the two actors, using a 9-point scale presented on the screen, anchored at one end by 1 (very dissimilar) and at the other end by 9 (very similar). Three similarity judgments were obtained for each pair: (1) How similar are the actors in terms of personality?, (2) Do these actors share a similar perspective on life?, and (3) Do these actors share similar values and attitudes? After all 15 pairs were presented, and the subjects' judgments obtained, the experiment ended, and subjects were debriefed.

Results

For all analyses, the three similarity judgments were averaged into a single judgment. Sex of subject did not show any significant main effects or interactions with other variables, and was not included in subsequent analyses.

The first questions concern the 10-point pairs. Does the interposition of a label and a boundary affect the degree of perceived similarity between members of a pair? To address these questions, the combined similarity measure was averaged over the three pairs appearing across category boundaries, and compared with the

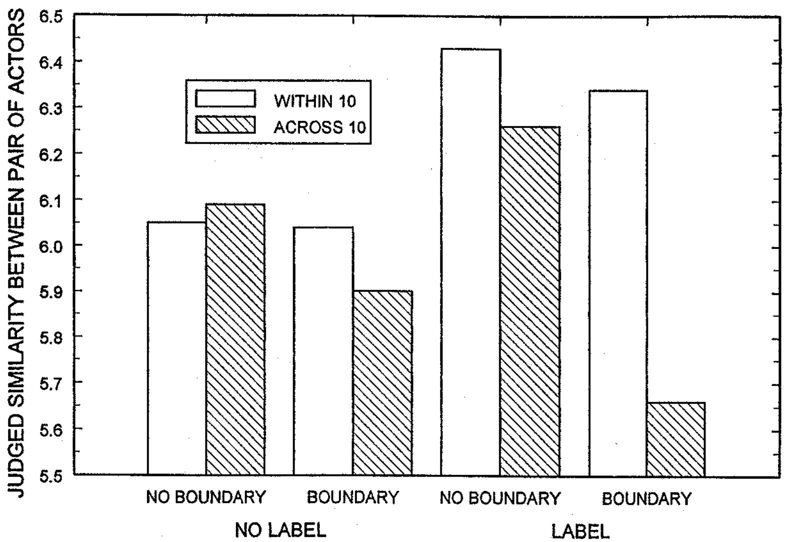

FIG. 3. Similarity judgments of pairs of actors, 10 points apart, by condition.

average over the four pairs appearing within a given category. The similarity judgments were subjected to a 2 × 2 × 2 ANOVA, with Condition (within vs across category markers) as a within-subjects factor, and Label (presence or absence of semantic label) and Boundary (presence or absence of heavy tick marks) as between-subjects factors. The findings are presented in Fig. 3. There is a significant effect of Condition [$F(1, 108) = 6.69$, $p < .02$], with pairs falling within the same category ($M = 6.21$) obtaining higher similarity ratings than pairs falling on opposite sides of a category boundary ($M = 5.97$). The difference between within- and across-category boundaries is accentuated with the addition of a label [Condition × Label $F(1, 108) = 4.25$, $p < .05$], with a greater difference (within minus across) when label is present (Mean difference = .43) than when it is absent (Mean difference = .05). The difference between within- and across-category boundaries is also increased with the presence of a visual boundary [Condition × Boundary $F(1, 108) = 3.67$, $p < .06$], with a greater difference (within minus across) when boundary is present (Mean difference = .41) than when it is absent (Mean difference = .06). The effects of Label and Boundary on Condition appear to be additive rather than multiplicative, since the interaction between Condition, Label, and Boundary is not significant [$F(1, 108) = 0.85$, ns).

The design of Experiment 3 allows us to distinguish between intercategory accentuation and intracategory minimization, a comparison lacking in Experiments 1 and 2. As indicated in Fig. 3, the no-boundary/no-label condition represents the baseline condition of similarity judgments for the two sets of pairs when no categorical information is available. The boundary present/label present

condition shows the strongest differences between the items existing within and across categories, and it is clear that the within items show greater similarity in that condition than they do in the baseline control condition, providing evidence for intracategory minimization. Moreover, the between items in the boundary present and label present condition show less similarity than in the baseline control condition, providing evidence for intercategory accentuation as well. This effect was tested using contrast weights of +1, 0, 0, −1 for the within/boundary and label, within/control, between/control, and between/boundary and label conditions, respectively, and was statistically significant [$F(1, 108) = 8.06$, $p < .01$].[7]

The second question generally concerns the effects of "reality" when pitted against categorization processes. It is possible to compare the perceived similarity of the 10-point pairs separated by labels/boundaries with the 15-point pairs that remain within the same category. Are objects "objectively" close together, but separated by arbitrary category markers, viewed as more similar to one another than are objects "objectively" further apart, but in the same category? The perceived average similarity of the three 10-point pairs across category markers was compared with that of the four 15-point pairs within the same category. Similarity judgments were subjected to a 2 × 2 × 2 ANOVA, with Condition (15-point within vs 10-point across category markers) as a within-subjects factor, and Label (presence or absence of semantic label) and Boundary (presence or absence of heavy tick marks) as between-subject factors. The findings are presented in Fig. 4. There is a significant effect of Condition (i.e., "reality") [$F(1, 108) = 7.85$, $p < .01$] with 10-point across pairs ($M = 5.97$) obtaining higher similarity ratings than 15-point within pairs ($M = 5.68$). The difference (within minus across) is diminished with the addition of an interposed label for the 10-point pairs [Condition × Label $F(1, 108) = 8.10$, $p < .01$], with a smaller difference between 10-point across minus 15-point within when label is present (Mean difference = −.01) than when it is absent (Mean difference = .60). Although the difference between within- and across-boundary conditions is also diminished with the presence of a visual boundary between the 10-point pairs, the difference is not significant [Condition × Boundary $F(1, 108) = 1.80$, $p < .20$].

It is worth noting, however, the differences in findings between the two extreme conditions with (1) no boundaries or label present and (2) both boundary and label present. The former condition shows the differences attributable to scale position alone, with no interposed category markers present. Under this condition the

[7] Since this contrast excludes the control conditions from the analysis, an additional test, which includes the control groups, was conducted. Using only the no label/no boundary and the label/boundary conditions, a 2 (condition: within vs across) × 2 (group: absence or presence of label/boundary) ANOVA, with repeated measures on the first factor, was conducted. There was both a significant effect of condition [$F(1, 55) = 7.66$, $p < .01$] and, most importantly, a significant interaction between condition and group [$F(1, 55) = 9.83$, $p < .005$]. That is, there was overall greater perceived similarity for within than across judgments, and this difference was significantly greater when labels and boundaries were present than when they were absent.

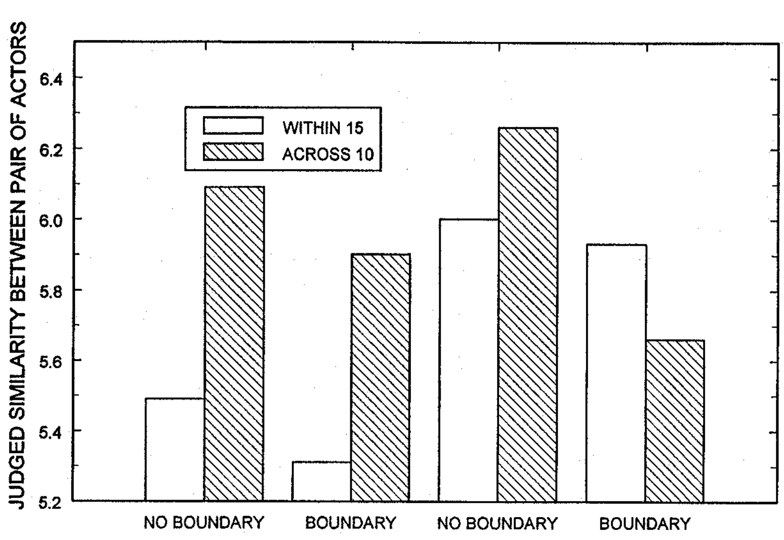

FIG. 4. Similarity judgments of pairs of actors, 10 and 15 points apart, by condition.

10-point pairs are seen as much more similar to one another ($M = 6.09$) than the 15-point pairs ($M = 5.49$). However, when both a label and boundary have been interposed between the members of the 10-point pair, perceived similarity is actually reversed, with members of the 15-point pairs seen as more similar ($M = 5.93$) than members of the 10-point pairs ($M = 5.66$). A contrast of the no label/no boundary condition with label and boundary condition showed highly significant differences [$F(1, 108) = 10.62, p < .01$].[8] Clearly, in this experiment, the effects of interposed category boundaries were successful in nullifying differences in perceived similarity due to actual scale position.

A third question concerns the number of category boundaries crossed. There were four pairs of actors located 35 percentile points apart. Two of those pairs crossed one major category marker, whereas two crossed two such category markers. Similarity judgments were subjected to a 2 × 2 × 2 ANOVA, with Condition (35-point across one category vs 35-point across two categories) as a within-subjects factor, and Label (presence or absence of label) and Boundary (presence or absence of heavy tick marks) as between-subject factors. Only one

[8] Again, an additional test including the control group was conducted. Using only the no label/no boundary and the label/boundary conditions, a 2 (condition: within vs across) × 2 (group: absence or presence of label/boundary) ANOVA, with repeated measures on the first factor, was conducted. Although there was no significant effect of condition [$F(1, 55) = 1.50$, ns], there was again, most importantly, a significant interaction between condition and group [$F(1, 55) = 10.73, p < .005$]. Although the pattern of means in the no label/no boundary conditions is reversed in the label/boundary conditions, there is no significant difference between the means in the latter group [$t(28) = 1.29$, ns).

significant effect was found: a Condition × Label interaction [$F(1, 108) = 6.27$, $p < .02$]. In the no-label condition, the average similarity of pairs crossing one category was lower ($M = 3.86$) than the average of pairs crossing two categories ($M = 4.30$), whereas in the label condition, this pattern was reversed (Means = 4.31 and 4.03, respectively). With the labels present, the pattern is as predicted, but the pattern in the no-label condition is unexpected. One possible explanation for the no-label pattern relates to the overall position of the two types of pairs on the continuum. One-category pairs are, on average, closer to the middle of the scale than are two-category pairs. For the one-category pairs, it is possible that having each actor on opposite sides of the midpoint makes them appear less similar than having one near the midpoint and one more polarized (as is the case for the two category pairs). Clearly, similarity judgments are influenced by a number of factors other than simply distance between the pairs, and absolute scale position seems to matter as well. Comparisons for the 35-point pairs should be interpreted with caution, but it is reassuring to note that the one significant finding showed that the effect of label was to increase the similarity of one-category pairs over two-category pairs, reversing the unexpected pattern found in the no-label conditions.

The final question concerns the effects of the judge's own scale position on similarity judgments. The large literature on social judgment, originating with Hovland and Sherif (1961), suggests that the ability to discriminate items on a continuum should decrease as the distance between the judge's own position and the items increases. This hypothesis was easily tested in the present study. Subjects were asked, after the similarity judgments were completed, to indicate their own position on the liberalism scale. For each subject, an absolute discrepancy score was obtained between the subject's own position and the average scale value for each of the seven 10-point pairs. That is, within each subject, for each 10-point pair, we computed the absolute distance between the position of the judge and the position of the stimulus pair (where the pair position was defined as the average of the two actors). Thus it was possible to calculate, within subjects, the correlation between the judge–pair discrepancy and the similarity rating for the pair. This correlation was calculated, across the seven pairs, transformed to Fisher's z, and then averaged across all subjects. The social judgment hypothesis predicts a positive correlation between the judge–pair discrepancy and the similarity rating—the greater the distance between judge and pair of actors, the greater the perceived similarity between the actors. In fact, the average correlation was $r = -.05$, and not significantly different from 0 [$t(108) = 1.44$, ns). Moreover, the magnitude of the correlation did not vary by condition when subjected to a 2 × 2 (Boundary × Label) ANOVA.

Discussion

Experiment 3 provided results generally consistent with the first two experiments and clarified the nature of the category boundary effects. The first question examined the effects of labels and boundaries on pairs 10 units apart, and found that both labels and boundaries significantly affected similarity judgments. Moreover, Experiment 3 showed evidence for both intracategory minimization and intercategory accentuation when categorized items (boundary and label present) were compared to a baseline control condition. Because a percentile scale is by definition continuous, and refers to the extremity of position in an absolute way (in comparison to others in the population), no substantive meaning is added by (*a*) labeling the different regions of the scale (i.e., "moderate liberal" or "liberal") or (*b*) providing vertical tick marks at the quartile locations (i.e., at 25, 50, and 75). Nonetheless, perceived similarity between the members of these pairs decreased as a function of both interposed labels and interposed boundaries. These effects were found between subjects, and are thus comparable to the findings in Study 1.

The second question examined the ability of categorical information to override the effects of "reality" by comparing 10-unit pairs whose members were located across the boundary markers with 15-unit pairs whose members were located within a bounded region. Under the no-boundaries/no-labels condition, where similarity is only determined by the "real" differences of 10 and 15 units, there is a large effect of reality in which the members of the 10-unit pairs are seen as much more similar to one another than are the members of the 15-unit pairs. However, as labels and boundaries are added, the effect is eventually reversed, so that 15-unit pairs remaining within a category boundary are viewed as more similar to one another than are members 10-unit pairs separated by both labels and boundaries. Although the general pattern is present for labels and boundaries, only the label effect reached significance in this comparison. Because these comparisons are within subject, but across items, they are comparable to the conditions run in Experiment 2.

The third question asked whether the perceived similarity between members of a pair decreased as the number of interposed category boundaries increased. Similarity ratings for pairs of actors differing by 35 units, but separated by either one or two category divisions, were compared. The results for this comparison generally confirmed the prediction for the effect of labels, with one-category pairs judged as more similar than two-category pairs when labels were applied to the continuum. Differences between the two pairs of items were present, however, even before labels and boundaries were applied. It is probably unwise to draw strong inferences from this comparison because the two types of items differed both on number of labels traversed and on relative position along the scale. The single-category pairs had to be located in a way that resulted in a higher average absolute distance between actor and midpoint of the scale than was true for the two-category pairs, which likely led to greater perceived similarity for the latter than the former.

Finally, the fourth question examined an explicit prediction from Hovland and Sherif's contrast principle in social judgment, and an implicit prediction from research on the perception of outgroup homogeneity: the greater the psychological distance between a judge and the objects of judgment, the greater the

perceived similarity among the objects. We found no evidence for this prediction, as tested by the within-subject correlation between the absolute distance from judge to objects and the perceived similarity of the objects. We have no ready explanation for the failure to support this prediction, but the finding is consistent with some literature on the perception of intragroup variability. There is a great deal of evidence in favor of greater perceived variability by ingroup than outgroup judges (Park, Judd, & Ryan, 1991), but the effect has only been obtained by group-level judgments, not by judgments of specific group members (e.g., Judd & Park, 1988). Since subjects in these experiments are judging the similarity of specific actors rather than the attributes of abstract categories, this research is consistent with some of the findings to date.

GENERAL DISCUSSION

Before discussing the implications of these findings, it would be useful to consider more generally the nature of assimilation and accentuation effects in categorization research. Assimilation refers to the homogenization of values within a category, usually in comparison to an uncategorized control. Assimilation, however, does not necessarily imply assimilation to central tendency, since it is possible to have within-category values assimilated to a value other than the arithmetic mean. For example, Lakoff (1987) describes prototypes based on a number of principles other than central tendency, including the use of "paragons" or idealized examples. In the Tajfel and Wilkes experiments, for example, the values of the "short" category could have been assimilated to that category's shortest line, and the values of the "long" category to that category's longest line—rather to the "average" line within each category. Although assimilation to central tendency was apparent in the Krueger and Clement studies, given subjects' clear knowledge of monthly variations in temperature, that is only one of several possible bases for assimilation.

Accentuation (or contrast) refers to the tendency to exaggerate differences between categories and has at least two possible meanings. First, it can mean that the central tendencies of the two categories are seen as further apart than is warranted (either by "reality" or an uncategorized control), or that perceived differences between objects at the boundaries of the categories are exaggerated. Although the Tajfel and Wilkes' research showed evidence for both forms of accentuation, it is possible to get the latter without the former. That is, it is possible to perceive veridically the means of adjacent categories, but exaggerate the perceived differences between the largest value of the smaller category and the smallest value of the larger category.

This brings us then to the relation between assimilation and accentuation: although accentuation can occur in the absence of assimilation (cf., the findings of Tajfel & Wilkes), it is difficult to imagine a case in which assimilation would not also produce accentuation of category boundaries.[9] As long as each value within

[9] It was possible to generate one (implausible) case where assimilation would not lead to

the category moves toward other values within the category, accentuation between the largest value of the smaller category and the smallest value of the larger category will occur. Indeed, the regression analysis by Krueger and Clement showed that a single principle, assimilation to central tendency, could account for category accentuation, and category boundary markers, as such, played no perceptible role in subjects' judgments.

The present experiments extend these earlier findings. Experiments 1 and 2 show clear effects of boundary position on similarity judgments, even when the position of the boundary varies (*a*) unpredictably for reasons irrelevant to the underlying continuum and (*b*) within subject. The design of Experiments 1 and 2 does not allow us to determine whether the greater similarity afforded to pairs within rather than across categories is evidence for (*a*) assimilation within category, (*b*) accentuation between categories, or (*c*) both. Although it seems unlikely that subjects are assimilating their judgments to the putative central tendencies of the categories "ideal," "acceptable," and "marginal," we cannot rule out this possibility with certainty. Whether the findings involve intracategory processes or not, Experiments 1 and 2 clearly show that subjects are attending to the category boundary positions. Experiment 1, using a between-subjects, within-item design, showed that the same two job applicants were judged as less similar when a boundary marker appeared between them, even though the category boundaries were said to be influenced by factors other than competence (number of job openings and number of applicants, which varied yearly). Experiment 2, conceptually replicated Experiment 1 with a within-subject, within-item design. This within-subject design is particularly significant, since nearly identical pairs of stimulus persons are judged differently by the same subjects as a function of changing boundary positions.

Experiment 3 does allow similarity judgments under categorized conditions to be compared to a baseline control condition, and those comparisons did show evidence that the presence of boundary markers produced both within-category assimilation and between-category accentuation. To our knowledge, this is the first demonstration of both effects within a single experiment. In this experiment, regions of the continuum were delineated both by the presence of verbal labels and by the use of tick markers to establish visual, nonsemantic boundaries. The effects of the visual boundaries is noteworthy, since they create barriers without the use of meaningful labels.

What theoretical mechanisms might account for these findings? Freyd (1983) has argued that the need to share concepts is central to meaningful communication, and at least in informal communication it is difficult to share concepts

accentuation of boundaries. Consider the Tajfel and Wilkes paradigm, and imagine that for the shorter category all values move to its longest line (No. 4) which itself does not move, and for all the values of the longer category to move to its shortest line which also does not move. Although this would not lead to accentuation between lines 4 and 5 (the lines defining the category boundaries) even though the within-category values are becoming more similar to each other, it is difficult to imagine why this might occur.

regarding an infinitely divisible continuum. For this reason in normal conversation we do not refer to a particular wavelength when describing color, but we use a limited number of category labels, with admittedly ambiguous boundaries [cf., Huttenlocher, Hedges, & Duncan's (1991) concept of boundary inexactness]. We speak of "orange" rather than 635 nm, or "heavy" rather than 100 kg. When these labels are frequently used, the label itself, and not the underlying physical value, becomes the psychological reality. Ashby's model of categorization emphasizes the mapping of the percept onto responses, with different categories mapping to different responses (Ashby & Lee, 1991). To the extent that subjects' implicit responses map more meaningfully onto such global categories as "ideal," "acceptable," or "liberal," rather than to the finer gradations associated with the underlying continuous scales, the boundaries that define these labels may become the functional perceptual region. Freyd's idea that communication and shared meaning help define the basis for categorization and Ashby's idea that objects in the same category elicit similar responses seem quite compatible to us, and together may be helpful in thinking about the categorization of social objects.

A recent experiment in the domain of color perception is relevant to our present findings. Goldstone (1995) presented, on a computer monitor, objects that varied in hue (e.g., individual letters ranging from red to red-violet) and asked subjects to reproduce the color of a particular object (e.g., the letter "L") on the monitor while the object was visible on the screen. The subject also was shown a second set of objects (e.g., individual numbers ranging from violet to red-violet), with the same instructions. Although a single letter (e.g., "L") would have the identical hue to a single number (e.g., "8"), the color reproduced for the letter tended to be redder than the presented letter, and the color reproduced for the number tended to be more violet than the presented number. In other words, the color reproduced for a specific object was influenced by the color associated with the category to which the object belonged. This result is particularly striking given that subjects were reproducing the color on the screen while the target object was displayed. Goldstone replicated these findings with objects drawn from classes of irregular geometric shapes.

If identical attributes (in this case, color) are judged differently because the objects are drawn from different categories, what constitutes "differentness" between categories? By one standard both numbers and letters are members of a single category—alphanumeric characters—yet by another standard, letters may themselves be subdivided further into vowels and consonants. Whether a set of objects is considered to belong to a single category or to multiple categories, can, to a large extent, be highly context dependent. One way to think about the imposition of categories on a continuous distribution is that the categories create many different objects out of a single object. In the first two experiments, a single dimension of competence is parsed into three domains of competence; in the third experiment the dimension of political liberalism is parsed into four regions or "objects." Different categories can be created then, not only by assigning different verbal labels, but by using other means for demarcating separation between adjacent regions—for example, the use of visual tick marks in Experiment 3.

These findings may be useful in interpreting the interesting results obtained on group attribution and outcome bias (Allison & Messick, 1985; Allison, Mackie, & Messick, in press). Consider two events: (1) 63% of city council members vote in favor of a new tax policy, which passes given majority rule, and (2) the same percentage vote by the same city council at a later point in time, fails to pass given a new requirement of a two-thirds majority to pass tax legislation. When asked how favorable the city council was to the new policy, respondents judged the city council as more favorable in the first condition (when it passed) than in the second condition (when it failed). This is parallel to our own findings where continuous data (the vote) are identical (more accurately, virtually identical), but the outcome falls into different categories: "pass" in one case and "fail" in the other case. One concern with some of the group attribution research has been that subjects may merely be readjusting their implicit anchor points on the scale when deciding whether a given vote is or is not favorable. Our findings, and particularly those of Goldstone, suggest that more than just scale adjustment may be responsible for giving greater weight to the categorical information. The group attribution research also raises interesting questions about the nature (and implications) of the labels themselves. "Pass–Fail" or "Win–Lose" are distinctions with more profound implications than "moderate liberal" or "liberal." Using the ideas of Freyd and of Ashby, would the magnitude of the category boundary effects depend upon the importance of the responses that map, implicitly or explicitly, onto the categorical distinctions?

The importance of assimilation and contrast effects in social perception has a long history. Hensley and Duval (1976) created a situation in which each subject's attitudes remained a constant distance from one set of group members (high similarity or ingroup condition), but systematically increased for another set of group members (low similarity or outgroup condition). Although the variability within ingroup and within outgroup remained constant and the position of the ingroup members to one another never actually changed, two very interesting consequences of increasing the distance between subject and outgroup members were found. As distance between subject and outgroup increased, the perceived similarity of self to ingroup also increased, and the perceived intragroup variability, within both ingroup and outgroup, decreased. That is, Hensley and Duval showed a strong increase in perceived similarity between self and other ingroup members by varying only the distance of self to outgroup, suggesting again that similarity judgments are highly dependent on context.

Contextual effects influencing perceived similarity of self to ingroup is also apparent in a study by Oakes, Turner, and Haslam (1991), who found that the judged typicality of a group member was dependent, not only on the behavior of the group member, but also on the degree to which that person's behavior was consistent with that of the group and different from outgroup members' behavior.

Similar to Hensley and Duval, the perceived similarity of a member to a group was also based on contrasts with outgroup members.

The problem for the experimenter interested in the effects of arbitrary boundaries is to provide category information to the subject that is truly perceived as meaningless. This is difficult, because we usually make the assumption that labels are provided for a reason—to add information. In the experiments reported here we attempted to create the perception of arbitrariness in a number of ways: (1) by describing the boundary positions as influenced by unpredictable factors, (2) by varying the position of the boundaries within subject, and (3) by using nonverbal, visual boundary markers. No single set of experiments can settle this issue definitively, and it is hoped that future research will be able to determine exactly how meaningless boundary markers can be made and still affect perceived similarity. Although the position of boundary markers along a continuum may be arbitrary, they have systematic effects on subjects' perception of similarity. In the real world of human groups, the placement of boundary markers is often arbitrary, but the effects are often quite systematic, and frequently much less benign than the examples used in this paper.

REFERENCES

Allison, S. T., & Messick, D. M. (1985). The group attribution error. *Journal of Experimental Social Psychology,* **21,** 563–579.

Allison, S. T., Mackie, D. M., & Messick, D. M. (1996). Outcome biases in social perception: Implications for dispositional inference, attitude change, stereotyping, and social behavior. In M. Zanna (Ed.), *Advances in experimental Social psychology* (Vol. 28, pp. 53–93). San Diego, CA: Academic Press.

Allport, G. W. (1954). *The nature of prejudice.* Reading, MA: Addison–Wesley.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General,* **120,** 150–172.

Brown, R. (1976). In memorial tribute to Eric Lenneberg. *Cognition* **4,** 125–153.

Freyd, J. J. (1983). Shareability: The social psychology of epistemology. *Cognitive Science,* **7,** 191–210.

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science,* **6,** 298–304.

Hensley, V., & Duval, S. (1976). Some perceptual determinants of perceived similarity, liking and correctness. *Journal of Personality and Social Psychology,* **34,** 159–168.

Hovland, M., & Sherif, C. I. (1961). *Social judgment.* New Haven, CT: Yale Univ. Press.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review,* **98,** 352–376.

Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology,* **54,** 778–788.

Krueger, J., & Clement, R. W. (1994). Memory-based judgments about multiple categories: A revision and extension of Tajfel's accentuation theory. *Journal of Personality and Social Psychology,* **67,** 35–47.

Lakoff, G. (1987). Cognitive models and prototype theory. In U. Nisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 63–100). New York: Cambridge Univ. Press.

Oakes, P. J., Turner, J. C., & Haslam, A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology,* **30,** 125–144.

Park, B., Judd, C. M., & Ryan, C. S. (1991). Social categorization and the representation of variability in formation. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 211–245). New York: Wiley.

Rosenhan, D. L. (1973). On being sane in insane places. *Science,* **179,** 250–258.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology,* **54,** 101–114.