Journal of Personality and Social Psychology
1994, Vol. 67, No. 1, 35-47

0022-3514/94/$3.00

Memory-Based Judgments About Multiple Categories: A Revision and Extension of Tajfel's Accentuation Theory

Joachim Krueger and Russell W. Clement

Accentuation theory states that the classification of stimuli produces encoding biases. Contrast effects enhance intercategory differences; assimilation effects enhance intracategory similarities. Do these biases affect the retrieval of stimuli distributed across many categories? The calendar superimposes arbitrary intermonth boundaries on day-to-day variations in temperature. In Experiment 1, Ss estimated the average temperatures of 48 days. Differences between estimates for 2 days belonging to neighboring months were greater (contrast) and differences between estimates for 2 days belonging to the same month were smaller than actual differences (assimilation). Experiment 2 showed that assimilation accounted for all categorization effects. When modified by assumptions from an exemplar model of category learning, accentuation theory accounts for the results. The relevance of these findings for social categorization is discussed.

Merely placing objects into distinct categories affects judgments about those objects. Objects belonging to the same class are more likely to be confused with each other (Stangor, Lynch, Duan, & Glass, 1992; Taylor, Fiske, Etcoff, & Ruderman, 1978) and to be recalled as a cluster (MacNamara, Hardy, & Hirtle, 1989). Often, only the category label is recalled (Park & Rothbart, 1982). People over- or underestimate spatial distances between stimuli depending on whether a category boundary falls between them (Newcombe & Liben, 1982), and they underestimate the degree to which the stimulus distributions of different categories overlap (Krueger & Rothbart, 1990). Categorization effects are intriguing because the act of categorization does not necessarily add information to the stimulus itself, but places a stimulus in a context. Categorization determines which stimuli belong to the same class and which do not.

In this study, we examine predictions derived from accentuation theory (Eiser & Stroebe, 1972; Tajfel, 1959, 1969). According to this theory, people's perceptions minimize the differences between stimuli falling into the same category (assimilation effect) and maximize the differences between stimuli falling into different categories (contrast effect). The standard design involves the presentation of stimuli that are graded along a quantitative dimension and divided by a categorical boundary falling in the center of that dimension (Tajfel & Wilkes, 1963). That is, accentuation effects have been demonstrated as encoding biases in cases of dichotomous categorization.

We first review theory and research on accentuation. Then, we ask whether predictions of accentuation theory can be applied to (a) cases where multiple mutually exclusive categories are superimposed on the stimulus array and (b) whether assimilation and contrast effects occur on the level of stimulus retrieval. To do this, we examine the implications of recent models of memory and categorization (Huttenlocher, Hedges, & Duncan, 1991; Smith & Zárate, 1992) for the classical accentuation paradigm. Next, we report two empirical tests of the modified accentuation theory, using judgments about average ambient temperatures as the target task. Finally, we highlight the similarities between this judgment task and social categorization.

A Review and Preview of Studies on Accentuation

At the root of accentuation theory are psychophysical studies on judgments about the quantitative properties of stimuli such as the spatial location of nonsense syllables (Campbell, 1956), the averages of series of numbers (Krueger, 1991; Krueger, Rothbart, & Sriram, 1989), or the length of lines (Tajfel & Wilkes, 1963). In social psychology, accentuation theory has provided a powerful paradigm for the study of stereotype formation and change (Haslam & Turner, 1992; Tajfel, 1969) and attitudinal judgment (Eiser, 1971; McGarty & Turner, 1992). The adaptation of accentuation theory to social categorization led to a spate of studies on intergroup differentiation. In that domain, the contrast effect has evaluative and descriptive variants. Evaluative differentiation usually emerges as in-group favoritism (Brewer, 1979), whereas descriptive differentiation involves the overestimation of between-groups differences on other, nonevaluative dimensions (Krueger, 1992). The assimilation effect emerges in perceptions of group homogeneity. Usually groups appear more homogeneous to nonmembers than to members (Linville, Fischer, & Salovey, 1989; Park & Rothbart, 1982; but see Simon & Brown, 1987).

The early studies on accentuation theory established a framework for research. Subsequent work has adopted and perpetu-

Joachim Krueger and Russell W. Clement, Department of Psychology, Brown University.

We thank Oliver John, Jill Portman, Myron Rothbart, Eliot Smith, Jack Wright, and two anonymous reviewers for helpful comments on a draft of this article. Bill Heindel provided invaluable statistical advice.

Correspondence concerning this article should be addressed to Joachim Krueger, Department of Psychology, Box 1853, Brown University, Providence, Rhode Island 02912. Electronic mail may be sent to jkrueger@brownvm.brown.edu

ated some of the original choices in experimental design, limiting the theoretical and methodological growth of the paradigm. We review some of these choices and then present two experiments intended to broaden the scope of theory, improve data analysis, and expand applicability of the accentuation paradigm. The review is focused on the number of categories used in classification, the kinds of judgments elicited, and the typical procedures in data analysis and their alternatives. We begin by reexamining Tajfel and Wilkes's (1963) classic psychophysical study.

Tajfel and Wilkes (1963) had subjects estimate the length of eight graded lines. In the experimental condition, the four shorter lines were labeled *A* and the four longer lines were labeled *B*. Thus, two categories were superimposed on the continuous variation in line length. In one of the control conditions, there were no category labels, and in the other, category labels were assigned at random and were thus uncorrelated with stimulus length. Difference scores were computed by subtracting the estimated length of one line from the estimate of an adjacent line. Difference scores involving the pair of lines straddling the category boundary were larger in the experimental than in the control conditions (contrast effect). However, the difference scores involving within-category pairs were not smaller in the experimental than in the control conditions (no assimilation effect). Subsequent studies have produced evidence for the assimilation effect (e.g., McGarty & Turner, 1992).

Number of Categories

Eiser and Stroebe (1972) noted that Tajfel's analysis was restricted "to the case of dichotomous classifications, but his predictions should apply equally well to situations with more than two classes" (p. 55). However, only one such study has been conducted (Wilder & Thompson, 1988). In studies on intergroup perception, subjects typically make judgments about one in-group and one out-group, but this two-category scenario is by no means exhaustive. Numerous social groupings involve multiple categories ordered along a continuously varying dimension. Educational and occupational roles are stratified by age. Political parties vary from Marxist to monarchist, with additional variation across countries. Social status ranges, in part, as a function of income, from lower to upper class. Finally, racial groups vary in the color of their skin and hair. Most important from the perspective of accentuation theory is the fact that the boundaries separating the groups are arbitrary social or cognitive–perceptual conventions. Yet, perceptions of the boundaries of social categories often contain a subjective component. The onset of middle age, for example, may in part depend on the age of the perceiver. The criterion of darkness of complexion for being a mestizo may vary with the darkness of the perceiver's skin. Whether an attitude is labeled *communist* depends on the political orientation of the labeler.

In our first test of the applicability of accentuation theory to a multiple-category scenario, we chose to eliminate individual differences in perceptions of category boundaries. Thus, we selected the calendar as a social convention that superimposes 12 mutually exclusive categories (months) on a continuously varying physical reality (ambient temperature). The use of multiple categories raises issues concerning the separation of contrast and assimilation effects, which we examine later in this article.

Kinds of Judgments

Making estimates about lines is responding to visual stimuli and quantifying one of their salient physical properties. This emphasis on judgments based on sensation and perception is part of the psychophysical heritage of accentuation theory, which has carried over into studies on attitudinal judgment, where the categorical context affects judgments of the extremity of statements (Dawes, Singer, & Lemons, 1972; Eiser, 1971). However, attitudinal or stereotype judgments not only show encoding but also retrieval biases. For example, judging the location on the left–right political continuum of the statement "Life and liberty are meaningless without the ownership of private property" (McGarty & Turner, 1992, p. 258) may be affected by the raters' recollection of his or her experiences with the political system and its prevalent philosophies. Relevant memory-based knowledge may be accessed and influence the appraisal of the stimulus.

The influence of memory on assimilation and contrast in categorization has barely been acknowledged within the accentuation paradigm. In this research, we suggest that judgments of mean temperatures for specific days, categorized by months, depend on a combination of three sources of information: memory traces for the temperature on the specific target days, memory for the temperatures on specific other but similar days, and generalized knowledge about climatic changes throughout the year. Two sets of theories with implications for accentuation effects in judgments made from memory are (a) prototype theories (e.g., Huttenlocher et al., 1991; Posner & Keele, 1968; Rosch, 1978) and (b) exemplar theories (e.g., Hintzman, 1986; Nosofsky, 1987; Smith & Zárate, 1992). We present the key assumptions of one typical theory of each type and explore their implications for retrieval biases in the accentuation paradigm.

A prototype theory. According to a recent prototype model (Huttenlocher et al., 1991), memory is hierarchical, containing knowledge of the central tendency (prototype) of the category, category boundaries, and individual exemplars. All three of these elements of knowledge are "inexact," involving error and uncertainty. The greater the uncertainty about a stimulus, the more the prototype is used as a default value for responding, resulting in a "shrinkage" of the dispersion of stimulus judgments. Hence, this prototype theory predicts assimilation but no contrast effects. Assimilation should emerge regardless of the number of categories that are judged concurrently, whereas according to accentuation theory, categorization effects occur only when at least two categories are judged.

An exemplar theory. According to a recent exemplar-based model (Smith & Zárate, 1992), separate memory traces for the central tendencies of categories are not necessary. When judging a target stimulus, perceivers instead "form a weighted average of the known values of all memory exemplars on that dimension," and "exemplars that are highly similar to the target will have the most weight" (Smith & Zárate, 1992, p. 9). This model predicts assimilation effects in judgments about categorized stimuli to the extent that attention to the category label

leads to (a) the recruitment of more exemplars of the same category and (b) greater weight placed on same-category than on other-category exemplars. When all target stimuli fall into the same category, the category label is less informative and receives less attention. Consequently, there should be no categorization effects in a single-category task. This prediction is consistent with accentuation theory but not with the prototype theory. It is tested in Experiment 2.

Measuring Accentuation

Category boundaries. The first of two important measurement issues is whether categorization effects occur only at the boundaries or across all stimuli. Tajfel and Wilkes (1963) compared judgments about stimuli that were adjacent on the physical continuum. This method limited the detection of contrast effects to exaggerated differentiation between stimuli at the category boundary. Later work has broadened the analysis and shown that boundary stimuli do not enjoy a special status. The overestimation of between-categories differences has been measured on the level of central tendencies (Eiser, 1971; Krueger & Rothbart, 1990; McGarty & Turner, 1992), suggesting that contrast effects resulted from perceptual contrasts involving judgments about all categorized stimuli. A reanalysis of Tajfel and Wilkes's data confirmed this idea (Krueger, 1992). All four of the short lines were judged to be shorter, and all four of the long lines were judged to be longer in the experimental than in the control conditions.[1]

In the two-category paradigm, analyzing contrast effects on the level of category means makes better use of the available data than restricting analyses to boundary stimuli. The use of multiple categories, however, reintroduces the question of whether the two types of analysis are equivalent. In the multiple-category case, a category falls between two neighboring categories so that both boundaries demarcate the beginning of another category rather than the beginning of an unoccupied range of the judgment scale. If both neighboring categories provide the context of generalized contrast effects, these two effects might cancel each other out. In the central category, a generalized contrast effect would involve decreased ratings in response to the higher category and increased ratings in response to the lower category. However, if boundary stimuli are indeed more sensitive to accentuation effects than are central stimuli, contrast effects may be observed in a multiple-category paradigm without displacement in the mean of the central category. The present work examines this possibility.

Effect size. The second measurement issue follows from Tajfel and Wilkes's (1963) statement that the typical experimental condition provides a "perfect [point-biserial] correlation between the classification and the physical attribute under judgement" (pp. 101–102). This correlation was considered important because it indicated perfect predictability of the categories and the stimuli they comprise. Note that statistically this statement is inaccurate. The maximum value of a point-biserial correlation is 1 only if there is variation between stimuli falling into different categories but no variation between stimuli falling into the same category. If eight stimuli fall into two categories of four, and if they differ by constant amounts from one to the next, then $r = .873$. In Tajfel and Wilkes's study, each line was about 5% longer than the next shorter one. Therefore, the between-stimuli differences (and thus the variance) were greater in the longer than in the shorter category. This difference in within-class variance further reduced the point-biserial correlation in the actual data ($r = .869$).

The correlation between categories and stimuli increases with decreasing variances within categories and increasing variances between categories. This is the logic of parametric tests of between-groups differences in the mean. The paradigm of accentuation theory may be conceptualized as a psychological analog of the analysis of variance (ANOVA), where the squared correlation between individual stimulus values and categories is the effect size. The goal of performing ANOVAs is to examine whether there is a significant effect of the independent variable (e.g., categorization). In the accentuation paradigm, these tests would always be significant and thus uninformative. The question is whether judgments in the experimental condition (systematic categorization) reveal a larger effect size than judgments in the control conditions (random or no categorization) or in the actual stimulus data. This was the case in Tajfel and Wilkes's (1963) study. The squared correlations were .80, .75, and .76 for the estimates in the experimental and control conditions and the actual data, respectively. In the present work, we hypothesize that the squared correlations derived from estimates of classified stimuli will exceed squared correlations derived from the actual properties of the stimuli. Because all correlations are bound to be large, we merely ask whether the direction of the difference is consistent with accentuation theory across conditions and experiments.

The issue of perceived predictability deserves a final comment. Psychologically, predictability obtained through intercategory differentiation is valuable. The quest for structure and certainty is often considered a basic human attribute, powered perhaps by both cognitive–perceptual and motivational forces (e.g., Kelly, 1955). Accentuation theory has finessed the question of why contrast and assimilation effects should occur when predictability is perfect even without perceptual distortion. As we have shown, predictability is below perfection in most cases and can be improved through accentuation effects. Moreover, it has been overlooked that predictability is asymmetrical. Eiser and Stroebe (1972) suggested that the direction of prediction does not matter (p. 52). Recall, however, that point-biserial correlations allow error-free predictions only for inductive inferences (from stimulus to category). Deductive inferences (from categories to stimuli) are subject to the statistical limitations noted above and should be regressive to be optimal. Given knowledge of the category but ignorance of specific stimulus characteristics, one can only predict the category mean and thus must accept within-category error (Einhorn, 1986).

---

[1] When short lines were judged, the mean of the estimates in the experimental conditions ($M = 17.29$ cm) was lower than in the control conditions ($M = 17.76$ cm) and the actual lengths ($M = 17.48$ cm). When long lines were judged, the means of the estimates in the experimental conditions ($M = 22.63$ cm) were higher than in the control conditions ($M = 22.20$ cm) and the actual lengths ($M = 21.23$ cm).

Overview and Hypotheses

The present research follows the paradigm of accentuation theory and extends it in two ways. First, as suggested by prototype and exemplar models, categorization should affect judgments made from memory. Second, categorization effects may not be limited to situations involving two categories. Therefore, we presented stimuli falling into multiple (12) categories (Experiment 1) and stimuli falling into 1 or 3 categories (Experiment 2).

Judgments about average temperature are well suited for a real-world test of categorization effects. In temperate climates, there is a gradual warming during one half of the year and a gradual cooling during the other half. People know these trends from experience, and they can estimate or predict the typical temperature for a specific day on the basis of specific recollections and generalized meteorological knowledge. The months of the calendar are categories with arbitrary boundaries superimposed on the gradual day-to-day changes in temperature. Both accentuation theory and the two memory models suggest that differences between estimates should be larger when the target days belong to different months than when they belong to the same month (tested in Experiment 1). Accentuation theory and the exemplar model, but not the prototype model, predict that judgments about stimuli falling onto a single category will not show assimilation effects (tested in Experiment 2).

In our own experience, categorization effects in temperature estimates are deceptively compelling. In Rhode Island, Septembers are cooler than Augusts (average high temperature = 75 and 80 °F, respectively). Suffering from the August heat, one of us (J. K.) expected temperatures to drop sharply with the arrival of September, rather than a gradual cooling across days and weeks. Similarly, forecasting frigid temperatures across the nation, Willard Scott, the noted weatherman of NBC's "Today" program exclaimed on January 27, 1994, "Geez, come on, February!" In light of the actual meteorological pattern, these expectations were unwarranted. It is conceivable that heat-induced discomfort motivated the expectation of cooling in September and that the icy January of 1994 motivated hopes for a sudden thaw in February. Accentuation theory as well as the memory models suggest, however, that such categorization effects are basic cognitive phenomena that occur even in the absence of wishful thinking.

Experiment 1

Subjects estimated the average high and low temperature for a series of 48 days. We predicted that relative to actual temperature changes, between-months changes would be overestimated (contrast effects) and within-months changes underestimated (assimilation effects). The overall trend in actual temperatures was a gradual warming between January and July and gradual cooling from July to January. There were no actual between-months differences when the global trends reverted from warming to cooling or from cooling to warming, and thus, no categorization effects were expected for these months. Because these global trends are quite familiar, the averages of the estimates for a given month were expected to be accurate.

Method

Subjects and procedure. One hundred seventy-seven undergraduate students at Brown University participated (58% women). Some received credit for introductory courses in psychology; others received a small payment. On their arrival in the laboratory, subjects were invited to participate in a number of different and unrelated tasks on social and physical judgment. One of these tasks, interspersed between others, involved judgments about temperature. Subjects worked individually in separate cubicles at Macintosh IIci personal computers in sessions that lasted about 1 hr.

Instructions were presented on the screen. It was explained that the experiment was concerned with people's ability to accurately estimate the average high and low temperature for a number of days across the calendar year. Information about the month of the stimulus day was presented lexically (e.g., "October 2") rather than numerically (e.g., "10/2"). On the screen, to the right of the presented day, there were two boxes in which subjects entered estimates of the average high and low temperature for the decade between 1981 and 1990. All estimates were made for the city of Providence, Rhode Island. Days were presented one at a time and in random order. Subjects typed their estimates in degrees Fahrenheit in the provided space.

Stimulus materials. Forty-eight days were selected, with four days per month. Two of these days (the within-month pair) fell close to the middle of the month and were eight days apart. The other two days were located close to the month boundaries and were eight days apart from the next stimulus day in the adjacent month (the between-months pair). The main analyses were focused on the difference scores derived from the within-month and between-months pairs. Note that the interday distance was constant for both types of pairs. The only difference was that a month boundary either did or did not fall between the days of a pair.

Actual temperatures of the years 1981 through 1990, as reported by the National Climatic Data Center, served as the criterion for accuracy (National Climatic Data Center, 1992). Figure 1 shows the average actual high and low temperature for the 48 target days. As would be expected in a semitemperate climate, the curve is sinusoidal, with a maximum in July and a minimum in January. Year-to-year and day-to-day fluctuations in temperature were considerable so that averaging across 10 years left noticeable residuals around the overall temperature trends.

Results

Preliminary analyses. Before testing the hypotheses derived from accentuation theory and the memory models, we assessed the correlational and mean-level accuracy of temperature estimates. Correlational accuracy is the correctness of perception of the relative ordering of the days in terms of their stimulus values irrespective of systematic over- or underestimation. It was assessed by correlating the estimated with the actual temperatures across the 48 target days. Correlations were computed for each subject separately and were averaged using Fisher's r-to-z-to-r transformation (McNemar, 1962). Correlational accuracy was high for high and for low temperatures (both mean rs = .91). Estimates of high and low temperatures were highly correlated as well (mean r = .97). Finally, the correlation between actual high and low temperatures was nearly perfect (r = .99).

Mean-level accuracy was assessed by computing for each subject and month the average estimated high and low temperature across the four target days. Averages of the actual temperature were computed in the same way. Results showed that high tem-

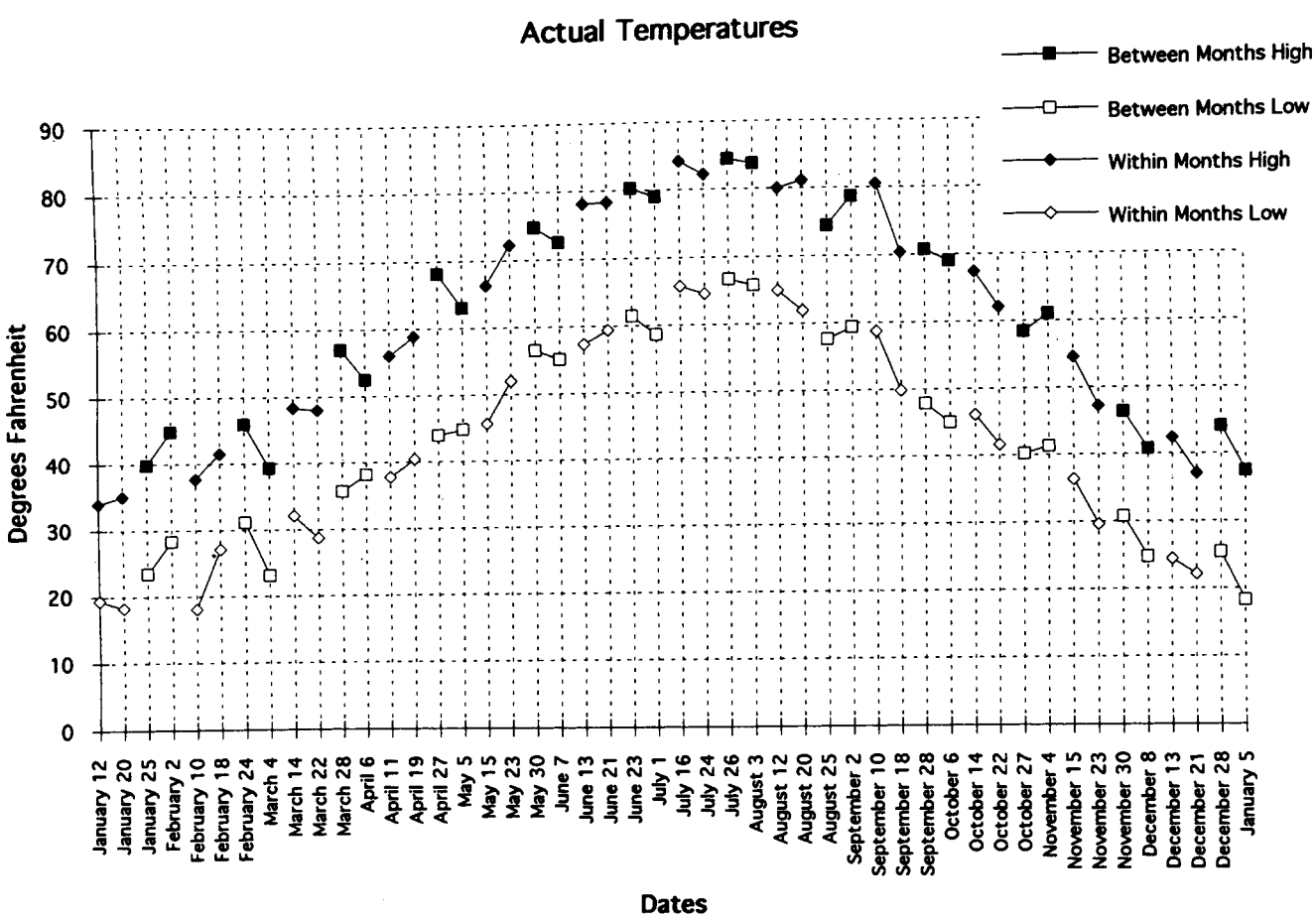

Figure 1. Actual high and low temperatures for each of 48 target days.

peratures were overestimated for four months (March, April, May, and August) and underestimated for three others (February, September, and October). There was also a consistent trend to overestimate the daily low temperature during the warming period (March through September), with an opposite result for February.[2]

Accentuation. Against this background of relative accuracy, accentuation effects were tested. The first analysis addressed the question of whether judgments exaggerated the predictability from categories to stimuli. That is, was the correlation between the group means of temperature estimates and months larger than the correlation between actual temperatures and months? Correlations were calculated separately for the first and the second half of the year because linear statistics were inappropriate for the sinusoid function of the year-round temperature changes. The warming period lasts from January 12 until July 1; the cooling period lasts from July 16 until January 5 (see Figure 1). Within the warming and the cooling periods, changes in actual and estimated temperatures approximated linearity, so that correlation coefficients could be computed. Their size showed that the assumption of linearity within periods was not violated. As is evident from Table 1, predictability was higher when computed with estimated temperatures (mean r^2 = .96) than when computed with actual temperatures (mean r^2 = .91). The differences were small but consistent across high and low temperatures and across both periods. The following analyses present more sensitive tests of accentuation effects.

The main hypothesis was that temperature estimates for a pair of days would differ more when a month boundary fell between them than when they belonged to the same month. Therefore, the estimate for the first day of a pair of target days (either between- or within-month) was subtracted from the estimate for the following target day. Because the warming trend during the first six months yielded positive difference scores and the

[2] The acceptable "familywise alpha error" (FWE; Keppel, 1982) was the probability that in at least 1 of the 12 comparisons, the null hypothesis would be rejected, although correct. With an FWE of .05, the modified Bonferroni test yielded a two-tailed probability of .004 for each of the 12 comparisons. Standard errors of the means were fairly homogeneous across months, ranging from .53 to .82. Confidence intervals for the means (ranging from 1.53 to 2.36) were calculated by multiplying the standard error of the mean temperature estimate for a given month with the z score of p = .002. The null hypothesis was rejected only if estimates differed by more than about 2 °F from actual temperatures. Thus, this test was conservative in accepting accuracy.

Table 1
Predictability of Daily Temperatures From Months: Squared Correlations Based on Actual Temperature Data and Human Estimates

Temperature	Period of the year	
	First half	Second half
Estimates with months		
High	.965	.962
Low	.964	.961
Actuals with months		
High	.901	.903
Low	.907	.946

cooling trend during the second six months yielded negative difference scores, two sets of difference scores were computed and then averaged across months and across the conditions of high and low temperature. Figure 2 shows the mean difference scores derived from temperature estimates and collapsed across high and low temperatures.

Difference scores were analyzed in a 2 (period: warming vs. cooling) × 2 (temperature: high vs. low) × 2 (categorization: between- vs. within-month) within-subjects ANOVA. The predicted interaction between categorization and period was significant, $F(1, 174) = 76.4$, $p < .001$. For the warming period, subjects estimated temperatures to increase more sharply between months than within months. For the cooling period, they estimated temperatures to decrease more sharply between months than within months. This effect was not qualified by an interaction with temperatures (high vs. low), $F(1, 174) = 1.3$, *ns*. The accentuation effect was highly consistent across subjects. Eighty-five percent of the subjects provided larger between-months difference scores than within-month difference scores ($z = 9.4$, $p < .0001$ by sign test). The main effect of period merely showed that subjects perceived warming during the first half and cooling during the second half of the year, $F(1, 174) = 231.4$, $p < .001$.

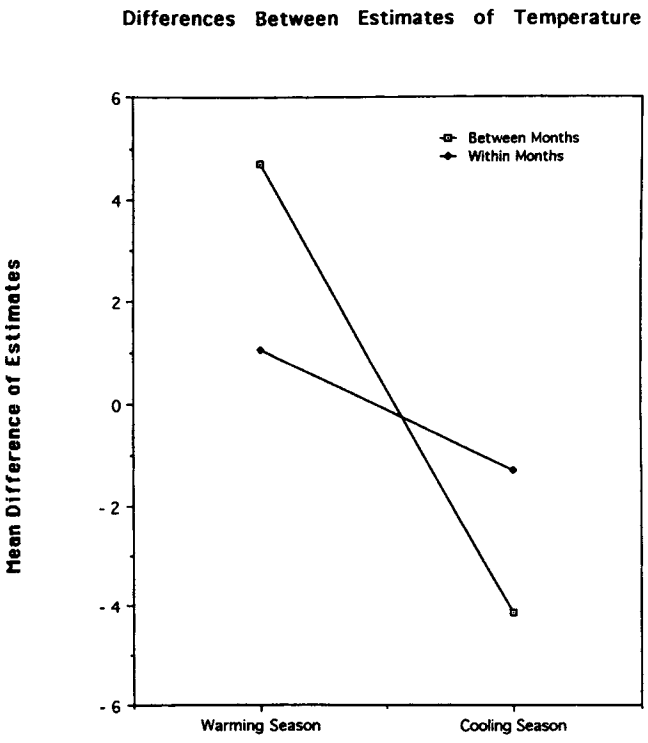

Figure 2. Mean differences between estimates of temperature.

The between-months accentuation effect could be interpreted as a categorization bias only if actual temperatures did not show the same effect. To ensure that actual temperatures did not change more sharply between months than within months (as we had found for the judgments), difference scores were computed in the same way as they were computed for the estimates. As should be expected, during the warming period, actual temperatures did not increase more sharply between months ($M = -0.15$ degrees) than within months ($M = 2.42$ degrees). Similarly, during the cooling period, actual temperatures did not decrease more sharply between months ($M = -2.25$) than within months ($M = -4.58$ degrees). In fact, as the means show, the between-months differences were slightly smaller than the within-month differences. These residual differences merely highlight the notorious volatility of New England weather. They were not tested for significance because they would vanish if temperatures were aggregated across a larger number of years.

Figure 3 shows that the estimates of the 48 mean high and low temperatures follow a step function, with sharper changes between than within months. There were no stepwise changes at the maximum and the minimum of the curve. This pattern was consistent with accentuation theory and the memory models. No accentuation would be expected in the absence of real differences.

Separate two-way ANOVAs were performed for each month to examine the stability of accentuation effects across the year. Both factors, between-months versus within-month differences and high versus low temperatures, were within-subject variables. Within-month differences were compared with their respective subsequent between-months differences. We predicted accentuation effects for all months except the ones in the middle of the summer and the winter, when overall warming and cooling trends are reversed. The results were strikingly consistent. Accentuation effects were significant for 8 months (all $ps < .001$).[3] As expected, the boundaries between June, July, and August and between December, January, and February did not yield accentuation effects. There were no main effects or interactions involving high versus low temperatures.

Contrast and assimilation effects. The preceding analysis revealed accentuation but permitted no conclusions about its constituent mechanisms. To test contrast and assimilation effects separately, differences between estimates of adjacent target days were compared with actual differences between the

[3] Again, an FWE of .05, yielding a probability of .004 per comparison, appeared acceptable for the rejection of the null hypothesis.

same days. If both effects contribute to accentuation, within-months differences should be underestimated and between-months differences should be overestimated. We calculated the percentage of subjects who over- or underestimated and compared them with the chance level of 50% (using sign tests). As predicted, most subjects overestimated between-months changes during the first (97%, $z = 12.4$, $p < .0001$) and the second half of the year (73%, $z = 6.1$, $p < .0001$). Similarly, most underestimated within-month changes during the first (68%, $z = 4.6$, $p < .0001$) and the second half of the year (90%, $z = 10.6$, $p < .0001$).

Discussion

Categorization effects specified by accentuation theory and memory models emerged in judgments about multiple categories. Differences between estimates of average ambient temperatures were larger when a month boundary fell between a pair of days than when both days belonged to the same month. Relative to actual changes, between-months changes were overestimated (contrast) and within-months changes were underestimated (assimilation). There were no categorization effects when the overall direction of the changes in actual temperatures reverted from cooling to warming or from warming to cooling.

The stability of categorization effects within seasons and across subjects suggests that these effects occur independently of motivational states. In the introductory examples, the predicted sudden cooling in September and warming in February could have been attributed to temperature-related discomfort during the preceding month. The present findings show that wishful thinking is not necessary to produce categorization effects in the predicted changes in temperatures. Finally, subjects were quite accurate on the level of average monthly temperatures. Means of the estimates for a given month correlated highly with the actual means, and most mean-level differences were smaller than 5 °F.

To clarify the implications of Experiment 1, three issues need to be addressed. First, do contrast and assimilation effects contribute independently to intercategory accentuation? Assimilation toward the month's central tendency and contrast away from the month's boundaries would yield the same result, namely, a reduction of dispersion of judgments within the month. The small number of target days per month did not permit independent tests of these two hypothetical mechanisms.

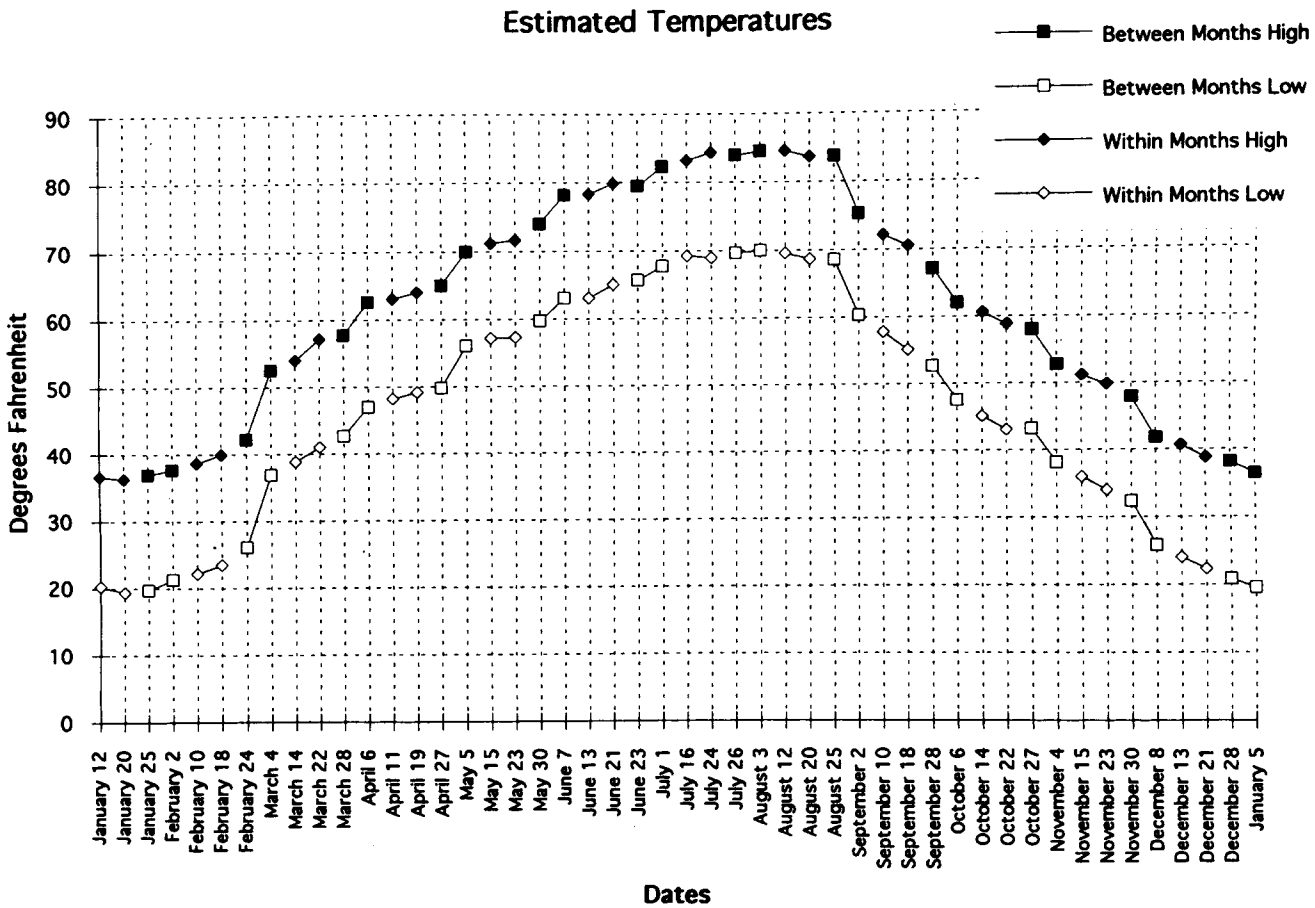

Figure 3. Estimated high and low temperatures for each of 48 target days.

Increasing the number of target days per month offers statistical possibilities for examining this question.

Second, accentuation theory and the two memory models predicted the same categorization effects. Under what conditions do these three approaches make different predictions? According to accentuation theory, categorization effects occur only when at least two categories are superimposed on an array of stimuli. The exemplar model is consistent with this idea, but the prototype model suggests that assimilation toward the central tendency occurs even for an array of stimuli classified into a single category.

Third, Experiment 1 had a single-condition design. Adding a condition with judgments about a single category will provide the needed control for the evaluation of accuracy. In laboratory studies of accentuation theory, control conditions either involve no category labels or labels that are uncorrelated with stimuli (see Tajfel & Wilkes, 1963). Both the reanalysis of Tajfel and Wilkes's data and the differences between the perceived and actual predictability (r^2) in Experiment 1 indicated that accentuation reduces accuracy. Experiment 2 addresses these issues.

Experiment 2

Procedures in Experiment 2 were similar to those in Experiment 1. However, to adequately test degrees of assimilation toward category prototypes, multiple exemplars per category were presented. If temperature estimates increase linearly within months, and if their slopes are smaller than the slopes of the actual changes, then contrast effects at the month boundaries will result as a byproduct of within-month assimilation. Recent work on accentuation theory did not find distortion specific to the boundary stimuli (Krueger, 1992; McGarty & Turner, 1992). To test the distortion hypothesis, 10 days were selected for each of three adjacent months. It was predicted that judgments for a pair of boundary days would not differ more than could be predicted on the basis of their respective within-month assimilation.

In Tajfel's (1969) work as well as in Experiment 1, stimuli fell into two or more categories. It is conceivable that a single category establishes a prototype and assimilation of stimuli toward this prototype. To test this idea, we asked another group of subjects to estimate temperatures for the target days of only one month. In line with accentuation theory and the exemplar model, we predicted that making concurrent estimates about days of more than one adjacent month is a necessary condition for categorization effects. In contrast, according to the prototype model, estimating temperatures for the days of one month is sufficient to elicit assimilation effects.

Theoretically, the increased (psychological) predictability in judgments of categorized stimuli may be associated with greater mean-level accuracy. That is, when subjects judge only the temperatures for days of a single month, accuracy may be fairly low. Early theory and research on accentuation indicated, however, that categorization effects are indeed biases that reduce accuracy. Thus, we predicted that accuracy would be greater in the single- than in the multiple-category condition.

Method

Subjects and procedure. One hundred eighty-eight undergraduate students at Brown University participated (59% women) in this experiment. All received credit for an introductory course in psychology. Procedures were the same as in Experiment 1, but this time, subjects estimated only the average high temperatures because categorization effects in Experiment 1 did not differ for estimates of high or low temperatures.

Materials and design. Ten days were selected for the months of March, April, and May. Target days were 3 days apart. Thus, there were two pairs of days bridging the between-months boundaries. The remaining 27 pairs fell within months. About half of the subjects were in the multiple category (MC) condition. These subjects made estimates for all 30 days. The other half was in the single category (SC) condition. These subjects were presented with only 10 target days, all of them belonging to March, April, or May.

Actual temperatures, as reported by the National Climatic Data Center (1992), served again as the criterion of accuracy. Because the gross trend within and between the three months was a linear increase, the average high temperature of each of the 10 days of each month was regressed on its serial position. The slopes represent the actual within-month variability. To test assimilation in estimates of temperature, estimates were regressed on the serial positions of the target days for each subject. In line with all three models, it was predicted that for each month, the average slope of the estimates would be smaller than the slope of the actual temperatures. Accentuation theory and the exemplar model predicted that the activation of multiple categories is necessary for the emergence of categorization effects. That is, differences in slope should appear in the MC condition but not in the SC condition.

Results and Discussion

The mean estimates for each day along with the regression lines for the actual temperatures are displayed in Figure 4, Panels A (MC condition) and B (SC condition).

Preliminary analyses showed again that temperature estimates were quite accurate. In the MC condition, the average within-subject correlation between the estimated and actual temperatures across the 30 days was high (mean $r = .89$). As in Experiment 1, subjects in the MC condition overestimated average high temperatures for March ($p < .001$) and April ($p < .01$) but not for May (*ns*). Consistent with the idea that categorization reduces accuracy, subjects in the SC condition overestimated average high temperatures only for March ($p < .05$).

Accentuation. As in Experiment 1, we first asked whether judging temperatures for several months would increase predictability of estimates from months. We computed the correlations between months (March = 1, April = 2, and May = 3) and the means of subjects' estimates for each day. Indeed, the squared correlations showed that daily temperature estimates were more predictable from months ($r^2 = .91$) in the MC condition than in the SC condition ($r^2 = .78$). The latter coefficient was virtually identical to the actual predictability of temperatures from months ($r^2 = .79$). Again, this difference is consistent with the idea that explicit multiple categorization leads to an exaggeration of predictability through accentuation and thus to a loss of accuracy.

For the second test of intercategory accentuation, the estimate for each target day in the MC condition was subtracted from the estimate for the following target day. The average of the two between-months difference scores ($M = 2.74$ degrees) was

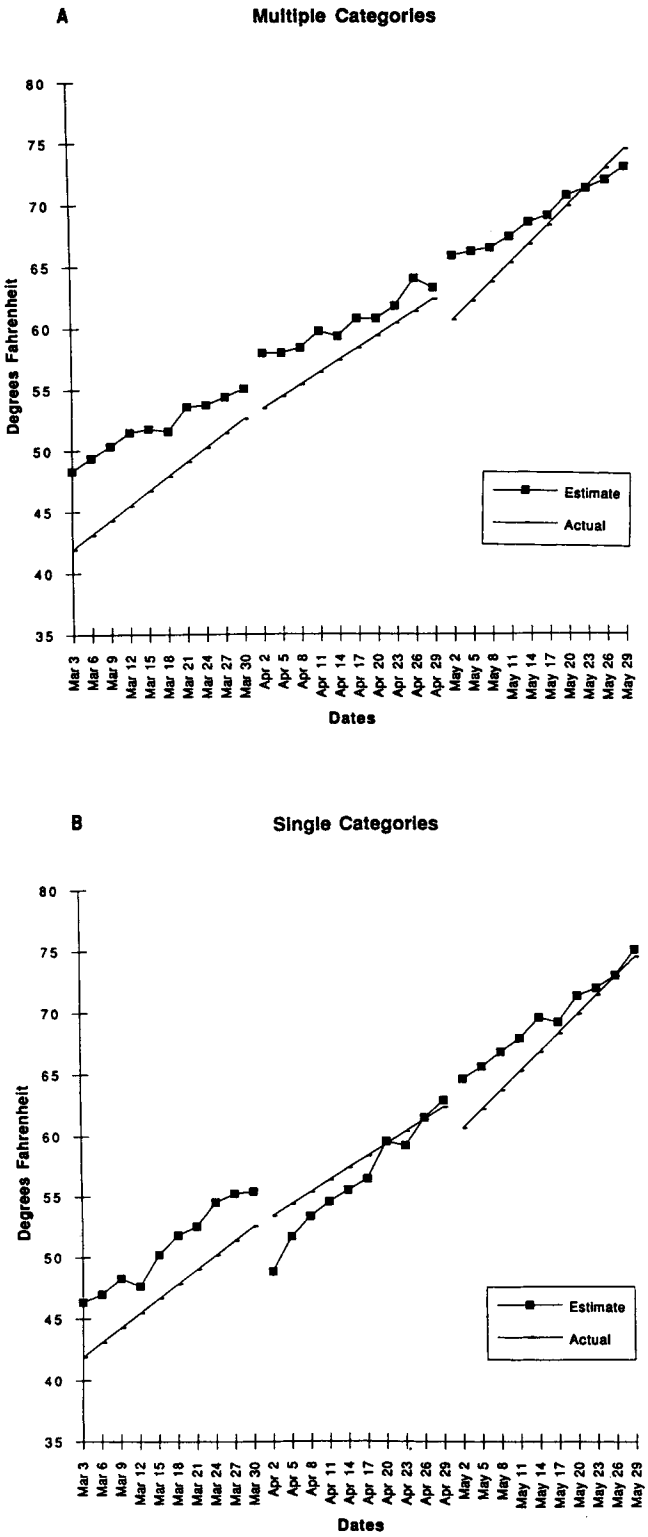

Figure 4. Average estimated and actual temperatures for the 30 target days in the multiple category condition (A) and the single category condition (B).

significantly greater than the average of the 27 within-months difference scores (M = 0.71 degrees), $t(98) = 3.6$, $p < .001$. In fact, none of the within-month differences was larger than any of the between-months differences ($p < .005$). In neither the SC condition nor the actual temperatures was the average of the between-months differences (Ms = −2.41 and 0.95, respectively) greater than the average of the within-months differences (Ms = 1.24 and 1.16, respectively).

Contrast versus assimilation effects. The original version of accentuation theory postulated additional contrast effects involving distorted perceptions of boundary stimuli. According to the memory models and more recent work on accentuation, assimilation effects should be sufficient to explain accentuation. To examine these different claims, assimilation and contrast effects needed to be tested independently. Therefore, we performed regression analyses separately for each subject, month, and condition. This procedure partitions the variance components due to subjects, items, and the interaction between subjects and items in repeated measures designs (Lorch & Myers, 1990). Only the days with serial positions 3–8 were entered in these analyses, leaving out the two days closest to each of the boundaries. The flatness of the slopes of these regression lines represented the degree of assimilation in the central section of the month.

Predicted values for the boundary stimuli were computed for each month and subject, using the regression parameters derived from the same subject. If the degree of assimilation did not vary with the distance from the prototype, the empirical estimates for the boundary days should not be different from the estimates predicted on the basis of within-month assimilation. However, according to the original accentuation hypothesis, the empirical estimates in the MC condition should be lower than predicted estimates at the upper boundary, and higher at the lower boundary. The evidence against this classical prediction of accentuation theory was strikingly consistent: None of the six paired t tests in the MC condition approached significance. Not surprisingly, predicted estimates for the boundary stimuli did not differ from actual estimates in the SC condition, either.

Assimilation without prototypes. We then examined the prediction of accentuation theory and the exemplar model that assimilation effects are limited to the MC condition versus the prediction of the prototype model that assimilation toward the central tendency of the category occurs regardless of the number of concurrently judged categories. To do this, we compared the slopes of the regression lines in the MC with the slopes in the SC condition and in the actual temperatures. In the "standard two stage" procedure (Feldman, 1988), within-subject beta weights (computed in Stage 1) are averaged and compared between groups (Stage 2). In the present analysis, all 10 estimates of each month were regressed on the serial positions. Accentuation theory and the exemplar model predicted that for each month, the mean beta would be smaller in the MC condition than in the SC condition or in the actual temperatures. The prototype model suggested that beta weights should be smaller in the SC condition than in the actual temperatures but not different from the MC condition. Table 2 shows the results.

An ANOVA on the within-subject beta weights revealed that, consistent with the prediction made by accentuation theory and

Table 2
Beta Weights for Actual and Estimated Changes in Temperature for Three Months

Condition	March	April	May
Multiple category	.51	.46	.61
Single category	.72	.81	.79
Actual temperatures	.74	.75	.95

the exemplar model, assimilation effects were stronger (i.e., yielding smaller beta weights) in the MC than in the SC condition, $F(1, 378) = 48.5$, $p < .001$. The degree of assimilation varied across months, $F(2, 378) = 5.4$, $p < .005$, but did not interact with the difference between conditions, $F(2, 378) = 2.3$, $p > .09$. Relative to the slope observed in the actual temperatures, assimilation in the MC condition was significant for each month (all ps $< .001$ by t test). However, in the SC condition, there was assimilation only for the month of May ($p < .001$). Mean beta weights for March and April did not differ from the actual changes in temperatures. Again, this analysis suggests that assimilation entails a loss of accuracy.

Taken together, the findings relevant to the three experimental questions suggest that neither the original accentuation theory nor the prototype model can account for all the data. On the one hand, as claimed by accentuation theory and the exemplar model, the concurrent presentation of at least two months was necessary for categorization effects. The finding that assimilation effects appear to be the main mechanism producing inter-category accentuation is consistent with the prototype model and the exemplar model. On the other hand, there was no evidence for contrast effects at the category boundaries. The data also cast doubt on the applicability of generalized contrast effects to multiple-category scenarios. Recall that a corollary of the generalized contrasts hypothesis was that with multiple categories, the effects of both neighboring categories on one central category should cancel each other out. Temperature estimates for the central month (April) showed as much assimilation, however, as did the two surrounding months.

General Discussion

The present research extended the paradigm of accentuation theory to memory-based judgments. Judgments reflected retrieval rather than encoding, and they involved either multiple or single categories, rather than the traditional two. As predicted by the theory, classification of stimuli into more than one category was necessary for accentuation effects to occur. These effects entailed (a) an increase in the predictability of (estimated) stimulus values from category labels and (b) a reduction in judgmental accuracy. The increase in predictability resulted from the minimization of within-category differences. The assimilation effects in the case of multiple categories were also consistent with a prototype model of memory, but the lack of assimilation in the case of a single category was not. Inconsistent with the original predictions of accentuation theory, there were no contrast effects either on the level of boundary stimuli or on the level of category means (i.e., no generalized contrast).

The results fit well with an exemplar model of judgment (Smith & Zárate, 1992). This model assumes that first, there are memory traces for individual events (e.g., the temperature on May 30). Second, the presentation of a target day does not only trigger the retrieval of the temperatures on that day over the last 10 years, but also the retrieval of the temperatures on similar days. Third, similarity, and thus the probability of retrieval, depends on the relative attention paid to the various stimulus attributes. Fourth, the judgment is an amalgamation (i.e., a weighted average) of the retrieved and relevant memory traces.

The critical feature of the exemplar model is the determination of stimulus similarity. A stimulus consisting of two attributes, month label and numerical position, recruits more related memory traces from the same month the more attention is focused on the month label. Attention to an attribute increases with increasing variability. If an attribute is invariant, as was the case in the SC condition of Experiment 2, attention is diminished. Inasmuch as the month label steers the retrieval of related days and their temperatures, assimilation effects occur. For example, the target day May 30 facilitates the retrieval of other days in May and inhibits retrieval of equally close days in June. As a consequence, the weighted average temperature estimate is biased toward the average temperature of May.

Given the good fit of the data with the exemplar model, should accentuation theory be abandoned? We hesitate to draw this conclusion and suggest instead that the present study illustrates how the applications of the two theories intersect rather than collide. Accentuation theory is a relatively narrow experimental paradigm, originally developed to account for encoding biases regarding categorized stimuli. Apart from Tajfel's (1969) occasional hints at a motive to maximize predictability, the theory has never supplied specific accounts of the psychological processes involved in accentuation effects. In contrast, the exemplar model is a broad-band cognitive theory of the representation in and retrieval from memory. This theory is insufficiently specific to guide the generation of hypotheses regarding social categorization. For example, the exemplar model did not inspire the hypothesis that multiple categories, when aligned along a continuous dimension, produce assimilation effects. Accentuation theory did.

The Arbitrariness of Categorization

The centerpiece of accentuation theory is the assumption that perception and judgment become biased even when inter-category boundaries are set arbitrarily. It is compatible with the exemplar model but does not necessarily follow from it. This assumption has had a significant influence on research and it has far-reaching consequences for social categorization and stereotyping. In the laboratory, it is easy to set category boundaries arbitrarily. Although Tajfel and Wilkes (1963) chose to draw the line between the short and the long stimuli between the two center stimuli, they did not have to. They could have placed the boundary between the third and the fourth line, for example. If set arbitrarily, boundaries can be moved within experiments and even within subjects. Krueger and Rothbart (1990) pre-

sented trait adjectives varying in favorability. Half of the traits were labeled *A* and half were labeled *B*, and these labels were highly correlated with favorability. Halfway through the experiment, the boundaries of one category shifted, creating (in some cases) overlap with the other category. As predicted by accentuation theory, identical traits were judged as more favorable when part of an overall desirable rather than an undesirable category (see also Davis-Stitt & Rothbart, 1993).

Accentuation experiments have shown that categorization biases occur even when boundaries are arbitrary. We suggest that a stronger statement is possible, namely, that many social boundaries are indeed arbitrary. This arbitrariness cannot be manipulated but can be inferred. We review some examples, starting with calendrical categories as a case in point and then return to the categorization of people into groups—the area on which accentuation theory had the greatest impact.

The arbitrary calendar. The month categories of the Gregorian calendar reflect salient celestial phenomena that are irrelevant for meteorological prediction. According to the Encyclopedia Britannica,

> The month is not suitable for determining the periods, for these are a solar, not a lunar, phenomenon. Periods vary in parts of the world—in tropical countries there are just the rainy and the dry periods, but elsewhere there are successions of wider ranges. (Encyclopedia Britannica, 1991, p. 433)

The Encyclopedia Britannica further explains that the main functions of the month are convenience, calculability, regulation of religious and civic activities, and finally its similarity to "the average menstrual period of women and also to the duration of cyclic behaviour in some marine creatures" (pp. 432–433). In other words, concerns with social structure, biological observation, and plain convenience lie at the heart of the calendar. Meteorological concerns are irrelevant. These characteristics make categorization by months more compelling than its alternatives. What these alternatives might be is not even clear. Solstices and equinoxes may divide the year into four seasons. Yet, many people are not sure about the precise dates of these events, and some prefer to group the seasons by months rather than by solar markers. We performed additional analyses (not reported in this article), without finding evidence for the idea that an alternative categorization by salient seasonal holidays (Memorial Day or Labor Day, for example) may evoke accentuation effects on a more superordinate category level.

Social categories. How arbitrary are the boundaries of social categories? The closest social analog to the present study is age. It is a continuous variable, along which both laypeople and professional lifespan psychologists superimpose multiple age-group categories (e.g., Kogan, 1979; Shanan & Kedar, 1979). The arbitrariness of the boundaries is apparent from the lack of consensus regarding their number and location and the preponderance of boundaries at the juncture or the middle of decades. Moreover, preferred boundaries change over time and new subcategories emerge. For example, Neugarten (1975) introduced a now popular distinction between the young-old (aged 65–70) and the old-old (70 and over). These categories, however arbitrary they may be, guide stereotypic expectations (Krueger & Heckhausen, 1993) as well as judgments about individuals (Krueger, Heckhausen, & Hundertmark, 1994). As in the present study, stereotype effects are stronger when raters evaluate multiple categories at the same time rather than individual categories in isolation (see a review by Kite & Johnson, 1988).

Accentuation theory has found greater use in studies on intergroup relations than in studies on age. Laboratory studies as well as some field studies (e.g., Sherif, 1961) have created groups using patently arbitrary criteria. One purpose of these studies was to model racial and ethnic conflict. Again, one need not assume that real groups are necessarily separated by boundaries that are somehow more objective or inevitable than the random separation of sophomores into blue and green laboratory groups. A strong claim of accentuation theory is that many categorical distinctions between social groups have come to seem "obvious" through mere repeated use rather than objective necessity.

Consider race. Whereas "physical anthropologists agree that race should be defined on the basis of population gene frequencies" (Yee, Fairchild, Weizmann, & Wyatt, 1993, p. 1133), they have failed to establish consensus on how many races there are. Counts vary from 3 to 37 (Yee et al., 1993). In practice, the criteria for racial membership often originate in the economic interests of dominant groups and are recorded in legalistic rather than scientific definitions (Harris & Kottak, 1963). Once in place, arbitrary criteria for "social races" can be perpetuated in a culture "by way of tradition and arbitrary practice" (Yee et al., 1993, p. 1134; see also Wagley, 1968). Hence, these definitions vary across cultures and with the groups to which they apply. In North Carolina before the Civil War, anyone with at least one Black ancestor five generations back was considered Black. In Brazil, anyone whose appearance is Caucasian to the slightest degree is considered White (Zuckerman, 1990). In the United States, anyone with at least one Native American ancestor four generations back is considered Native American. Similarly, the number of defined race groups is variable. The Republic of South Africa imposes multiple race categories (Blacks, Whites, and *Coloreds*, i.e., people of mixed ancestry; Fredrickson, 1981). Not surprisingly, anthropologists question the utility of such legal or commonsense definitions of race for scientific work because within-group variability far exceeds between-groups differences on most dimensions of interest (Latter, 1980). In social cognition, however, these definitions control perception, memory, and judgments of people. Accentuation theory helps understand how.

Of the three great social categorization schemes—age, race, and sex—might sex be the only truly qualitative distinction? Developmental geneticist Fausto-Sterling (1992, 1993) argued that "biologically speaking, there are many gradations running from female to male; and depending on how one calls the shots, one can argue that along that spectrum lie at least five sexes—and perhaps even more" (Fausto-Sterling, 1993, p. 21). A person's permanent sexual classification is not dictated by nature but required by law. Every newborn must be registered as either male or female. For ambiguous cases, modern medicine supplies hormonal and surgical procedures, which "disambiguate" and assimilate. In the middle ages, some hermaphrodites were buried alive (Fausto-Sterling, 1993).

How Social Is Categorization?

The juxtaposition of temperature categories, small laboratory groups, and age, race, and sex categories may be bewildering, but it illustrates the capacity of accentuation theory to incorporate a broad range of phenomena. This universalist philosophy dates back to the New Look, which championed a social approach to perception. The contrast effect, first observed in estimates of the size of coins (Bruner & Goodman, 1947), was replicated by Tajfel and Wilkes (1963) under conditions devoid of socially conditioned value judgments. To label superimposed categories merely *A* or *B* (Krueger et al., 1989; Tajfel & Wilkes, 1963) is to render categorization essentially meaningless. Recently, there have been a number of attempts to increase the effect size of accentuation by imbuing minimal, nonsocial categorizations with social meaning (Krueger, 1991; Krueger et al., 1989; McGarty & Turner, 1992). These attempts have been unsuccessful, thus further blurring the distinction between social and generic categorization.

Depending on one's point of view, it is possible to construe individual studies as examples of either social or generic categorization. Tajfel and Wilkes's (1963) thought experiment of treating nationality (Swedish vs. Italian) as a categorical variable superimposed on continuous variations in height is illustrative. Krueger (1991) reported accentuation effects in judgments about categorized numbers that ostensibly represented psychometric test scores of two fictitious groups of students (Group A and Group B). From a social psychological perspective, these examples conveyed information about people; from a general cognitive perspective, there were numbers to be averaged and to be remembered. In the same vein, the present experiments may give rise to alternative readings. From a social psychological perspective, it is clear that structuring time by months is a social convention (see the Encyclopedia Britannica, 1991). Calendars are expressions of the cultures that produce them. From a general cognitive perspective, the stimuli (or rather "memory probes") were just numbers representing a physical reality.

The universalist position downplays the differences between social and nonsocial categorization (Ostrom, 1984). Although it is undeniable that social categorization is unique in a number of ways, the differences between social and nonsocial categorization may have been overstated. The categorization phenomena studied by cognitive and social psychologists are quite similar conceptually but different methodologically. The "mental comparison" approach to categorization effects, which dominates cognitive research, is focused on reaction time measures in comparative judgments about paired stimuli rather than on quantitative judgments about individual stimuli (see Cech, Shoben, & Love, 1990; or Maki, 1981, for examples of this method). Ironically, accentuation theory itself holds a clue of why the mental-comparison approach and accentuation theory seem rather different. The two research areas use distinct terminologies and thrive in different scientific communities. In other words, categorical labels are superimposed on the graded theoretical and procedural differences. This, in part, may be reason enough to perceive social and general categorization as more distinct than they really are.

References

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive–motivational analysis. *Psychological Bulletin, 86,* 307–324.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42,* 33–44.

Campbell, D. T. (1956). Enhancement of contrast as composite habit. *Journal of Abnormal and Social Psychology, 53,* 350–355.

Cech, C. G., Shoben, E. J., & Love, M. (1990). Multiple effects in judgments of magnitude. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1142–1152.

Davis-Stitt, C., & Rothbart, M. (1993). *Accentuation effects in categories with variable boundaries.* Unpublished manuscript, University of Oregon.

Dawes, R. M., Singer, D., & Lemons, F. (1972). An experimental analysis of the contrast effect and its implications for intergroup communication and the indirect assessment of attitude. *Journal of Personality and Social Psychology, 21,* 281–295.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50,* 387–395.

Eiser, J. R. (1971). Enhancement of contrast in the absolute judgment of attitude judgments. *Journal of Personality and Social Psychology, 17,* 1–10.

Eiser, J. R., & Stroebe, W. (1972). *Categorization and social judgment.* San Diego, CA: Academic Press.

Encyclopedia Britannica, Inc. (1991). Calendar. In P. W. Goetz (Ed.), *The New Encyclopedia Britannica* (Vol. 15, pp. 432–449). Chicago: University of Chicago Press.

Fausto-Sterling, A. (1992). *Myths of gender: Biological theories about women and men* (2nd ed.). New York: Basic Books.

Fausto-Sterling, A. (1993). The five sexes. *The Sciences,* March/April, 20–24.

Feldman, H. A. (1988). Families of lines: Random effects in linear regression analysis. *Journal of Applied Physiology, 64,* 1721–1732.

Fredrickson, G. M. (1981). *White supremacy: A comparative study in American and South African history.* New York: Oxford University Press.

Harris, M., & Kottak, C. P. (1963). The structural significance of Brazilian racial categories. *Sociologica, 25,* 203–209.

Haslam, S. A., & Turner, J. C. (1992). Context-dependent variation in social stereotyping 2: The relationship between frame of reference, self-categorization and accentuation. *European Journal of Social Psychology, 22,* 251–278.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

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

Kelly, G. A. (1955). *The psychology of personal constructs* (Vol. 1). New York: Norton.

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed). Englewood Cliffs, NJ: Prentice Hall.

Kite, M. E., & Johnson, B. T. (1988). Attitudes toward older and younger adults. *Psychology and Aging, 3,* 233–244.

Kogan, N. (1979). A study of age categorization. *Journal of Gerontology, 34,* 358–367.

Krueger, J. (1991). Accentuation effects and illusory change in exemplar-based category learning. *European Journal of Social Psychology, 21,* 37–48.

Krueger, J. (1992). On the overestimation of between-group differences. *European Review of Social Psychology, 3,* 31–56.

Krueger, J., & Heckhausen, J. (1993). Personality development across

the adult life span: Subjective conceptions versus cross-sectional contrasts. *Journal of Gerontology, 48,* 100–108.

Krueger, J., Heckhausen, J., & Hundertmark, J. (1994). *Perceiving middle-aged adults: Effects of stereotype-congruent and incongruent information.* Manuscript submitted for publication.

Krueger, J., & Rothbart, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality and Social Psychology, 59,* 651–663.

Krueger, J., Rothbart, M., & Sriram, N. (1989). Category learning and change: Differences in sensitivity to information that enhances or reduces intercategory distinctions. *Journal of Personality and Social Psychology, 56,* 866–875.

Latter, B. (1980). Genetic differences within and between populations of the major human subgroups. *American Naturalist, 116,* 220–237.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57,* 165–188.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149–157.

MacNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 211–227.

Maki, R. H. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 15–32.

McGarty, C., & Turner, J. C. (1992). The effects of categorization on social judgement. *British Journal of Social Psychology, 31,* 253–268.

McNemar, Q. (1962). *Psychological statistics* (3rd ed.). New York: Wiley.

National Climatic Data Center. (1992). *National climatic data for Providence, RI, 1981–1990.* Ashville, North Carolina: U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service.

Neugarten, B. L. (1975). The future of the young-old. *Gerontologist, 15,* 4–9.

Newcombe, N., & Liben, L. S. (1982). Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology, 34,* 46–58.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–108.

Ostrom, T. M. (1984). The sovereignty of social cognition. In R. S. Wyer Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–38). Hillsdale, NJ: Erlbaum.

Park, B., & Rothbart, M. (1982). Perceptions of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology, 42,* 1051–1068.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Rosch, E. H. (1978). Principles of categorization. In E. H. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Shanan, J., & Kedar, H. S. (1979). Phenomenological structuring of the adult life span as a function of age and sex. *International Journal of Aging and Human Development, 14,* 343–357.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation. The Robbers Cave experiment.* Norman: University of Oklahama.

Simon, B., & Brown, R. (1987). Perceived intragroup homogeneity in minority and majority contexts. *Journal of Personality and Social Psychology, 53,* 703–711.

Smith, E. R., & Zárate, M. A., (1992). Exemplar-based model of social judgment. *Psychological Review, 99,* 3–21

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62,* 207–218.

Tajfel, H. (1959). Quantitative judgement in social perception. *British Journal of Psychology, 50,* 16–29.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues, 25,* 79–97.

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

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36,* 778–793.

Wagley, C. (1968). *The Latin American tradition: Essays on the unity and diversity of Latin American culture.* New York: Columbia University Press.

Wilder, D. A., & Thompson, J. E. (1988). Assimilation and contrast effects in the judgments of groups. *Journal of Personality and Social Psychology, 54,* 62–73.

Yee, A. H., Fairchild, H. H., Weizmann, F., & Wyatt, G. E. (1993). Addressing psychology's problems with race. *American Psychologist, 48,* 1132–1140.

Zuckerman, M. (1990). Some dubious premises in research and theory on racial differences. *American Psychologist, 45,* 1297–1303.

Received April 28, 1993
Revision received December 6, 1993
Accepted December 27, 1993 ■