

# EUROPEAN REVIEW OF SOCIAL PSYCHOLOGY

## VOLUME 3

*Edited by*

WOLFGANG STROEBE  
*University of Tübingen*  
and

MILES HEWSTONE  
*University of Bristol*

JOHN WILEY & SONS

Chichester • New York • Brisbane • Toronto • Singapore

## Chapter 2

# On the Overestimation of Between-group Differences

Joachim Krueger  
*Brown University*

## ABSTRACT

The overestimation of between-group differences is a central characteristic of social stereotyping. The present review focuses on exemplar-based category learning. Evidence is presented for the assumptions that: (a) contrast effects need to be distinguished from accentuation effects; (b) both effects are general cognitive-perceptual phenomena; and (c) they affect stereotype formation as well as stereotype change. The implications of the overestimation of between-group differences for person perception and conflict resolution are discussed.

Long before the experimental study of intergroup relations and stereotyping had even begun, Freud characterized the effects of social grouping on attitudes:

Of two neighbouring towns each is the other's most jealous rival; every little canton looks down upon the others with contempt. Closely related races keep one another at arm's length; the South German cannot endure the North German, the Englishman casts every kind of aspersion upon the Scot, the Spaniard despises the Portuguese. We are no longer astonished that greater differences should lead to almost unbearable repugnance, such as the Gallic people feel for the German, the Aryan for the Semite, and the white races for the colored. (Freud, 1921/1959, p. 33.)

*European Review of Social Psychology, Volume 3* Edited by Wolfgang Stroebe and Miles Hewstone  
© 1992 John Wiley & Sons Ltd

Although the examples Freud adduced in order to construct his psychology of the masses today seem dated and perhaps too drastic, his views ring true considering the vehement and persistent intergroup hostilities that the media continuously deliver from around the world. Embedded in Freud's analysis are three components of intergroup attitudes that have consistently emerged from the experimental research during the past three decades: ingroup favoritism; the perception of group homogeneity; and the emphasis on intergroup differences. However, the search for parsimonious explanations amenable to experimental testing has shifted from deep-seated conflicts or emotional needs to perceptual and cognitive processes of categorization.

Gordon Allport (1954) noted the critical role of categorization in the formation of stereotypes. He concluded that stereotyping results from normal thought and is inevitable.

The mind must think with the aid of categories. Once formed, categories are the basis for normal thought and judgement. We cannot possibly avoid this process. Orderly living depends on it (Allport, 1954, p. 19).

Tajfel (1969) presented a cognitive theory in which he suggested that the three major phenomena of stereotyping are facets of intergroup differentiation. They emerge when people are categorized into mutually exclusive groups, and when the observer is a member of one of them.

*Ingroup favoritism* is the tendency to value one's own group and its members more highly than groups one does not belong to. Sociologists have described this tendency as "ethnocentrism" in natural (e.g. ethnic or racial) groups (Sumner, 1906). Psychologists prefer the term "ingroup favoritism", which is applicable to any kind of group, category, or aggregate of people. Studies in the minimal-group paradigm have demonstrated that ingroup favoritism arises even when group members do not interact personally, are not competing for scarce resources with the other group, and do not possess a long-term identification with their own group. Billig and Tajfel (1973) arranged extraordinarily minimal conditions for group formation and found ingroup bias even when subjects knew they were randomly divided into groups (see Diehl, 1990, for a recent review).

The perception of *group homogeneity* is popularly known as overgeneralization. It describes the belief that members of a group are similar to one another, and that referring to them by their group label conveys all that is necessary to know about them. In most cases, outgroups are perceived to be more homogeneous than ingroups (Judd & Park, 1988; Linville, Fischer & Salovey, 1989). Park and Rothbart (1982), for example, found that even groups who had a considerable amount of contact with each other (e.g. men and women) attributed fewer stereotypic and more counterstereotypic traits to themselves than to the other group. However, under certain conditions

perceptions of group homogeneity are stronger with respect to ingroups rather than outgroups. Simon and collaborators found perceptions of ingroup homogeneity when subjects were members of a minority (Simon & Brown, 1987) or when they strongly favored the ingroup over the outgroup (Simon & Pettigrew, 1990).

Freud's concept of the "narcissism of small differences" (Rothbart & John, in press) pointed to the role of *intergroup differences* in stereotyping. Freud conceived of intergroup differences as a stimulus activating "sediments of feelings of aversion and hostility" (Freud, 1921/1959, p. 33), i.e. ingroup favoritism, and he expected the magnitude of aversion to increase with the magnitude of the difference. Perceived intergroup differences and ingroup favoritism are two facets of Tajfel's interclass effect, while the perception of group homogeneity is the intraclass effect.

In this chapter, I suggest that: (a) ingroup favoritism and perceptions of intergroup differences are conceptually different; and (b) their frequent co-occurrence in social perception is an important antecedent of social conflict. In the review of the experimental evidence, I will focus on perceived intergroup differences and on cognitive distortions leading to the overestimation of such differences. In brief, I present two theses. First, people perceive greater intergroup differences than there really are. Second, intergroup conflict is facilitated when overestimated differences become associated with ingroup favoritism.

Stereotyping starts with perceptions of readily identifiable personal characteristics and proceeds to inferred characteristics and expected events. D.T. Campbell showed in numerous studies that when there are real intergroup differences, perceivers are likely to enlarge these "grains of truth" (Campbell, 1967, p. 823). Prejudiced subjects accentuated differences in skin color between Blacks and Whites more than did non-prejudiced subjects (Secord, Bevan & Katz, 1956). Similarly, White, but not Black students overestimated actual differences in academic performance between the two racial groups (Clarke & Campbell, 1955; also Klineberg, 1944).<sup>1</sup> Where racial or other salient cues are absent, presumed attitudinal dissimilarity often serves as a means for group formation and intergroup perception (Rokeach, 1960). Dawes, Singer and Lemons (1972) hypothesized that a person "exaggerates the discrepancy between his own attitudes and the attitudes represented by opinion statements endorsed by people with opposing views" (p. 281). In pilot testing, attitude statements that were pro- or anti-US involvement in Vietnam were selected, and items varied in the degree of extremity. Then students with "hawkish" or "dovish" attitudes about the war were presented with a series of pairs of items. Each pair

<sup>1</sup> For a noteworthy exception to the "kernel-of-truth hypothesis," see McCauley & Stitt (1978, Experiment 3).

consisted of a hawkish and a dovish statement. When asked to select the more extreme items from each of a pair, there was a contrast effect.

The hawks ranked all of the dove extreme statements as more extreme than any of the hawk statements, while the doves ranked all the hawk statements as more extreme than any of the dove extreme statements (p. 287).

In these exemplary studies, a classification was superimposed on continuously varying attributes. There were real differences between the groups and these differences were exaggerated in judgments. In line with the traditional "kernel-of-truth hypothesis," one might conclude that real differences lie at the heart of many stereotypes. Work in the minimal-group paradigm has convincingly shown, however, that real differences are not even necessary for the perception of differences. Howard and Rothbart (1980) found illusory perceptions of differences between ingroups and outgroups where in fact no differences existed. In two experiments, subjects were arbitrarily categorized according to a patently irrelevant psychological attribute (under- and overestimators of dots). Nevertheless, subjects developed expectancies that ingroup members differed favorably from outgroup members, thus exhibiting both ingroup favoritism and intergroup separation. Moreover, differential expectancies biased memory toward better recall for favorable ingroup attributes and unfavorable outgroup attributes. Such expectancies even arise when subjects are categorized in an unabashedly random fashion. By way of a lottery, Locksley, Ortiz and Hepburn (1980) ostensibly classified groups of students into Phi and Gamma. In fact, all subjects drew Phi lots. The evident randomness of the procedure and the choice of Greek letters virtually precluded differential assumptions about the similarity of ingroups and outgroups; the often-used classifications based on alleged rating differences in dot estimation tasks or aesthetic preference do not. Yet, Locksley *et al.*'s subjects (1980, Experiments 1 and 2) allocated more token chips to Phi members than to Gamma members, and they expected desirable personality attributes to be more characteristic of ingroupers than of outgroupers.

Erroneous intergroup distinctions may be perpetuated by selective information search following initial categorization. Wilder and Allen (1978) arbitrarily distinguished between admirers of Klee or Kandinsky paintings and asked them to fill out an attitude inventory. Then subjects rank-ordered their preferences about information they might receive about other participants. Information emphasizing their attitudinal similarity with ingroup members or dissimilarity with outgroup members was preferred over information emphasizing dissimilarity with ingroup members or similarity with outgroup members.

Implicit in the minimal-group paradigm is the assumption that differential

perceptions of groups facilitate social conflict, and unwarranted or exaggerated differences in group evaluation may serve to justify hostility and discrimination. In typical studies, participants allocate more monetary rewards to their fellow members than to outgroupers (Tajfel, 1970). One reason for such discrimination may be the belief that positive reinforcement is more motivating for ingroup members than for outgroup members. Rothbart and Hallmark (1988) designed a simulation of an international crisis, where subjects were asked to take the perspectives of citizens of either of two imaginary countries, Takonnia and Navalnia. Both countries were described as arguing over the access to an area rich in mineral deposits and being locked in a destructive arms race. Both countries had developed a new generation of arms ready to be deployed. Subjects then selected a policy option from a list of alternatives that varied from conciliatory (Country X will unilaterally stop its production of new weapons and cut back existing forces by 20%, with the expectation that Country Y would make the same cutback in its forces) to coercive (Country X will build up its new weapons and threaten to use them unless Country Y cuts back on its new weapons). In trying to discourage the opponent country from deploying its weapons, subjects favored relatively coercive means, while they believed that their own country was more likely to respond favorably to gentle persuasion. In real life, the media may contribute heavily to perceptions of intergroup threats. Winter (1987) coded politicians' speeches for power-related statements. Subsequent content-analyses of partisan newspapers revealed that printed reports of speeches accentuated the power motive of the opponent politician, thus magnifying the threat and justifying rigorous responses.

To summarize, the (exaggerated) perceptions of intergroup differences are built into social perception. Differences are observed where none exist; and when they do exist, they are apt to be overestimated.

## STEREOTYPE LEARNING AND CHANGE

The core of this chapter is a review of a series of studies recently conducted at the University of Oregon. The goal was to explore exemplar-based stereotype formation and change and, specifically, to determine the minimum conditions leading to systematic overestimation of between-group differences. The experimental work rested on four assumptions:

1. Stereotypes can be acquired in a purely *exemplar-based* fashion.
2. Judgments about target groups are affected by the implicit or explicit context of comparison categories.
3. No *ingroup-outgroup distinctions* are necessary for the overestimation of between-group differences.

4. When stereotype-related attributes are interval-scaled, categories can be represented by intuitively estimated *means*.

### Exemplar-based Category Learning

Two distinct sources of information contribute to the formation and change of stereotypes. Exemplar-based information refers to information specifically associated with an individual, as, for example, the statement "Ms Buxley has blue eyes." In contrast, abstraction-based information refers to attributes that are in some way characteristic of the entire category, as in the statement "Swedes are blond." Park and Hastie (1987) pointed out that these two sources of information are usually intertwined. The first assumption of the present review was that differentiated intergroup perceptions can emerge from the observation of group members alone. That is, generalized prior beliefs about between-group differences are not a necessary condition for stereotyping. This assumption is not trivial, because such generalized beliefs have been shown to affect the encoding of observed individual characteristics. Biernat, Manis and Nelson (1991), for example, presented subjects with pictures of men and women. Ratings of height were made on an objective scale (feet and inches) or on a subjective scale (tall vs. short). On both measures, the specific men were judged to be taller than the women, although there was no real difference in height in this sample of target individuals. Subjects evidently relied on the accurate stereotype that in the population men are on the average taller than women.

In an experiment which has deservedly become a citation classic, Tajfel and Wilkes (1963) showed that intercategory differences are overestimated even when prior beliefs about the categories are effectively ruled out. Subjects estimated the lengths of eight graded lines. When the four shorter lines were labeled A and the four longer lines were labeled B, there was an interclass effect. The length of the longest A line was underestimated and the length of the shortest B line was overestimated. There was no such effect when the categorization did not covary with length or when there was no categorization (for similar studies see Lilli & Lehner, 1971; Petzold, 1990).

### Stereotyping in Context

The second assumption was that stereotyping is context-dependent. All three facets of intergroup differentiation, ingroup favoritism, perceptions of group homogeneity, and perceptions of intergroup differences, are contextual. Perceptions of groups cannot develop without (at least implicit) comparisons with other relevant groups. In Katz and Braly's (1933) early work on ethnic

stereotypes, subjects simply selected "characteristic" traits for a number of groups. One can hardly know what subjects had in mind when they judged a trait to be characteristic for a group. It is possible, however, that judgments were based on implicit between-group comparisons. The stereotypic musicality of Blacks probably did not mean that the majority or all Blacks were believed to be musical, but rather that relatively more Blacks than, for example, Japanese or White Americans, were believed to be musical.

McCauley and Stitt's (1978) diagnostic-ratio measure of stereotyping takes this possibility into account. They defined stereotypic traits as traits that are believed to be *relatively* more prevalent in a target group than in humanity. That is, beliefs about humanity provide the base rate against which specific groups are compared. This point is important, because it shows that mere estimates of trait prevalence in a given group are uninformative. Someone who believes, for example, that 60% of all Chinese are industrious, may appear to hold a positive attitude toward this group if the percentage is evaluated against the midpoint of the scale. If, however, the same person believes that 80% of all people (or some other reference group) are industrious, her attitude toward the Chinese appears relatively unfavorable.

### Beyond Ingroup-outgroup Distinctions

The third assumption of the present paradigm was that distinctions between ingroups and outgroups are not necessary for the overestimation of between-group differences. There is no doubt that such distinctions are important. In within-subjects designs, where observers judge both ingroups and outgroups, the salience of own-group membership may contribute to the differential treatment of outgroups. It is even conceivable that, when no ingroup judgments are made, observers spontaneously invoke the prevalence of a trait in their own group as an easily accessible base rate against which they judge outgroups. However, contrast effects have long been known to occur in a variety of domains independent of ingroup-outgroup distinctions. Two examples from the diverse areas of weight estimation and human attraction may illustrate the ubiquity of contrast effects.

Early psychophysical studies on weight-lifting (e.g. Rogers, 1941) demonstrated the role of judgmental anchors. When target weights were paired with heavy anchor weights they appeared lighter than when they were judged individually. Such contrast effects were reliably obtained unless the anchor weights were only slightly heavier than the heaviest target weights (Sherif, Taub & Hovland, 1958), and when both weights were perceived as part of the same task environment. In an ingenious experiment, Brown (1953) asked subjects to help him prepare the testing session by picking up the tray with the target weights and taking it to a desk. Although the weight of the tray was identical to that of the official anchor weight, subjects did

not construe this activity as part of the judgmental context. Hence, their ratings of the test stimuli were not influenced by the lifting of the tray.

Kenrick and Gutierrez (1980) and Kenrick, Gutierrez, and Goldberg (1989) examined the psychophysics of romantic love. They showed that judgments of the pleasantness of human faces and nude bodies were influenced by the pleasantness of previously viewed individuals. For example (Kenrick, Gutierrez, & Goldberg, 1989, Experiment 1), they had subjects first rate the attractiveness of *Playboy* or *Penthouse* centerfolds and then rate the attractiveness of average-looking women. Both male and female subjects rated the average-looking women as less attractive when they had viewed the centerfolds first.

Amster (1964) interpreted such contrast effects in the light of adaptation-level theory (Helson, 1964), suggesting that the totality of available exemplars determine the rater's adaptation level, and that individual stimuli are not evaluated in terms of absolute standards, but depending on the degree to which they deviate from the adaptation level. Amster's study was germane for the area of stereotype formation because she used lists of words varying in favorability. Unfavorable words were judged less pleasant when presented within an evaluatively heterogeneous list than when presented within an all-negative list. She did not, however, classify words *a priori* into favorable and unfavorable categories.

Accentuation theory, as already discussed with the example of Tajfel and Wilkes's (1963) study, went beyond contrast effects that can be attributed to the presence or absence of extreme anchor stimuli. The mere categorization of stimuli into two distinct classes is sufficient to produce an accentuation of perceived category boundaries. Unlike Amster's (1964) research, Tajfel and Wilkes (1963) presented the full range of stimuli (short lines and long lines) in the experimental and the control conditions. The superimposition of a dichotomous category was sufficient for the perceptual sharpening of the boundaries.

Recently, Davis-Stitt (1989) replicated the interclass effect with socially relevant stimuli, and she introduced two intriguing modifications to the traditional design. Subjects were given seven descriptions of fictitious job applicants who varied in the degree of suitability for employment. Ostensibly, applicants had already been evaluated by professional recruiters and had received summary scores. Superimposed on these interval-scaled scores were the three discrete categories "ideal," "acceptable," and "marginal." The first modification was that the instructions stressed the arbitrariness of the category boundaries. It was pointed out that depending on the relative numbers of applicants and jobs, the ranges of scores considered ideal, acceptable, or marginal could vary from one year to the next. In fact, the boundaries were varied between subjects. For example, the fifth applicant was categorized as ideal for some subjects and as acceptable for others. The

second modification concerned the dependent variable. Instead of judging single stimuli, subjects rated the similarity of pairs of descriptions. The interclass effect was predicted to emerge as a decrease in rated similarity when a pair straddled category boundaries. Of the seven pairs that appeared both within and between categories, six showed the predicted discrepancy, and three differed significantly. Only for one pair were similarity ratings slightly higher when the two descriptions fell into different categories.

In sum, these studies suggest that in contexts unrelated to ingroup-outgroup distinctions, the very existence of category boundaries has powerful effects. Boundaries may be patently arbitrary and, in order to sharpen these boundaries perceptually, observers need not locate themselves on either side of the fence.

### The Role of Central Tendencies

The fourth assumption was that category accentuation need not be reduced to biased perceptions of boundary stimuli. Rather, exemplar-based knowledge about categories or groups may form distributions that can be represented by central tendencies and variances. Whether intercategory differences are overestimated or not, can be assessed on the level of the true means of individual stimulus ratings as well as on the level of intuitively estimated means. Consider Tajfel and Wilkes's (1963) study. To quantify the interclass effect, these authors focused on the differences between judgments about lines that were adjacent on the continuous scale of length. In the experimental condition, but not in the control conditions, the difference between the two border lines was greater than the differences between any pair of lines belonging to the same category. Instead of merely comparing adjacent stimuli within conditions and then comparing the differences between conditions (Tajfel & Wilkes's procedure), one might ask whether each individual line was rated more extremely (shorter or longer) in the experimental than in the control conditions. Indeed, the presented results (p. 106) revealed that in the experimental conditions, *all* lines were rated as more extreme than in the control conditions. To reanalyze the original data, for each line, estimated lengths in the control conditions were subtracted from estimates in the experimental conditions. Table 2.1 shows the results.

Clearly, boundary stimuli had no privileged place in the interclass effect. The magnitude of judgmental displacement because of categorization was fairly homogeneous across stimuli. If anything, there was a tendency for ratings of outlying lines to be *more* displaced than boundary stimuli. Overall, the data suggest that interclass effects may be captured more appropriately by category means. In a further reanalysis, the average length of the short lines was subtracted from the average length of the long lines. The resulting

**Table 2.1** Mean differences in centimeters between judgments of length of lines in the experimental and control conditions

Stimuli	1	2	3	4	5	6	7	8
Classification minus control	-0.7	-0.5	-0.4	-0.4	0.4	0.6	0.3	0.5

Data were computed from Table 2 in Tajfel & Wilkes (1963, p. 106)

difference was greater in the experimental condition ( $M = 5.34$ ) than in the control conditions ( $M = 4.44$ ) or for the true length ( $M = 3.75$ ). In the control conditions, the difference should not have been significantly different from the true difference, and it probably was not. Because standard deviations were not provided in the original publication, it is impossible to test the significance of the interclass effect in the experimental conditions.

More recent tests of the interclass effect have abandoned the focus on boundary stimuli. In Eiser's (1971) study, 64 attitude statements were rated with respect to their restrictiveness with respect to drug use. When the 32 more permissive statements were attributed to one newspaper and the 32 more restrictive statements to another, *average* ratings for the two sets of statements were more polarized than when the statements were presented in the absence of newspaper names. Similarly, McGarty and Penny (1988) found that, on average, judgments of political statements were polarized when the radical and the reactionary statements were attributed to different authors.

It should be kept in mind that these category averages were computed from subjects' individual responses, and not estimated by the subjects themselves. Lewis (1990) introduced an intriguing method of studying mental models of category means. He presented 30 Chernoff faces (simple drawings containing the basic facial features). Subjects learned to classify them into the groups A and B, depending on the width of the nose. Noses of A faces were four units wider than the neutral face, and B faces were four units narrower. In a reconstruction task, Lewis presented a neutral Chernoff face and asked subjects to modify the facial features until they represented the "typical" A face. This procedure was then repeated for the reconstruction of the "typical" B face. As predicted, the reconstructed faces showed more extreme nose widths than the true group averages.

In addition to the collection of individual stimulus ratings and subsequent computation of group averages, it seems reasonable also to ask subjects to estimate the mean of the ratings they had made. After all, when stereotypes about groups are formed, people cannot make use of actuarial averaging procedures, keeping track of all group members they have met and the impressions they have formed about them. Rather, to represent the group

as a whole, they must rely on the information they can retrieve from memory and intuitively aggregate this information into a single score. This procedure is particularly compelling when the number of available exemplars is large.

## THE CATEGORY-LEARNING EXPERIMENTS

### Overview

The paradigm underlying the experiments reviewed in this section builds on the four assumptions discussed above: category learning was exemplar-based; involved two categories as a judgmental context; did not presuppose ingroup-outgroup distinctions; and estimated means were taken as adequate category representations. In all experiments, subjects were presented with two intermixed series of stimuli (trait terms in some experiments, numbers in others). Before a stimulus was shown on screen, a letter (A or B) briefly appeared as a category identifier, telling subjects to which of the two categories the following stimulus would belong. The stimulus itself was then shown for a few seconds. When it disappeared, subjects typed in the category-identifying letter and made a quantitative judgment about the stimulus. When numerical stimuli were used, they simply typed in the number they had seen, and when trait adjectives were used, they entered a favorability rating using a scale from 0 (extremely unfavorable) to 100 (extremely favorable). Several times during the procedure, subjects also estimated the mean for each category. They were informed about this demand at the beginning of the sessions and assured that the experiments were no tests of their mathematical acumen. It was made clear that the task of simultaneously keeping track of two cumulative means was challenging but not impossible.

The stimuli formed two relatively flat but approximately normal distributions. In most cases (an exception will be discussed later) the distributions bordered one another but did not overlap. One distribution (the *focal category*) comprised stimuli from the middle range of the scale, that is, more or less neutral trait terms or medium-large numbers. The other category (the *contextual category*) comprised either stimuli from the low range or from the high range of values. That is, traits were either very unfavorable or very favorable, and numbers were either smaller or larger than focal numbers. Thus, by holding the distribution of the focal category (middle range) constant, and by varying the distribution of the contextual category, it was possible to explore context-dependent shifts in the perception of the focal stimuli and estimates of the focal mean.

In line with the general assumption that intercategory differences are overestimated, two specific biases were predicted: contrast effects and

accentuation effects (Krueger & Rothbart, 1990). *Contrast effects* refer to context-induced distortions in the perception or the evaluation of individual stimuli. Neutral traits may appear more favorable when presented along with negative traits than when presented along with positive traits (Amster, 1964). If subjects base estimates of category means on distorted individual ratings, even accurate intuitive averaging will result in contrasted mean estimates. *Accentuation effects* refer to any additional exaggeration of between-category differences that cannot be traced to biased perceptions of individual stimuli. Above and beyond perceptual distortions, subjects may give greater weight to those exemplars that sharpen intercategory distinctions than to those that blur such distinctions. That is, accentuation effects occur when mean estimates are more biased by the context than the true averages based on individual judgments.

Each experiment consisted of two phases, a *category-learning phase* and a *category-change phase*. In the category-learning phase, the basic statistical parameters (mean, variance, range) of the two distributions remained stable, while the number of stimuli was gradually increased. In the category-change phase, one or more of these parameters were changed in the focal category. The nature of these changes and their effects will be discussed below.

### Trait Experiments

In one experiment (Krueger & Rothbart, 1990, Experiment 3), subjects were presented with 96 trait adjectives, which had been selected from Goldberg's (1973) list of 1710 personality descriptors. Purportedly, the traits described members of two fictitious groups, A and B. Goldberg's (1973) normative social desirability data served to group traits so that in phase 1 there was a neutral focal category (ranging from "extravagant" to "worldly"), and either a negative contextual category (ranging from "dishonest" to "boisterous") or a positive contextual category (ranging from "obliging" to "honest"). Based on the presented trait information, subjects rated the likeability of each member, and they estimated both group means at the end of phase 1 (that is, after a total of 48 traits had been shown). In line with the predicted contrast effect, the average favorability judgments in the neutral focal group were significantly higher ( $M = 52.25$ ) when the contextual group was negative than when it was positive ( $M = 44.08$ ). Moreover, intergroup differences were accentuated as the average estimated mean for the focal group was even further displaced from the negative group ( $M = 64.49$ ) than the computed average of the individual ratings. When the contextual group was positive, the average estimated mean ( $M = 45.47$ ) did not differ from the computed mean.

Phase 2 of this experiment was designed to test the idea that certain changes in the distribution of the focal category may lead to intercategory

accentuation. Specifically, it was expected that extreme exemplars carry greater weight in mean estimation than moderate exemplars. Focal stimuli that were different from the contextual stimuli were conceived of as extreme, and focal stimuli that were similar to the contextual category were conceived of as moderate. Subjects were presented with another set of 48 traits. The *a priori* distribution of the contextual category remained constant, with traits ranging either from "inhumane" to "cocky," or from "verbal" to "dependable." In the focal group, however, the variance and the range became larger, with traits ranging now from "insincere" to "sincere." The resulting range of favorability in the focal group completely encompassed the range in the contextual groups.

At the end of phase 2, contrast effects were as strong as after phase 1. When the contextual group was negative, averaged favorability judgments were higher ( $M = 53.30$ ) than when it was positive ( $M = 45.80$ ). The accentuation of intergroup differences on the level of estimated group means was even more dramatic. Averaged estimated means in the focal group were even further displaced from the contextual groups than they were after phase 1 ( $M$ , with negative context = 66.73;  $M$ , with positive context = 41.95). The observed accentuation effects revealed illusory mean change. The mere symmetrical increase of variance in one group was sufficient to induce an increase in perceived intercategory differences. Estimated group means (but not averaged individual ratings) shifted away from the contextual group, although there was no true change.

What are the psychological processes underlying the phenomenon that extreme exemplars "carry greater weight" in mean estimation? One possibility is that there are differences in memory accessibility. Focal traits eliciting the same affective response as contextual traits may have been mistakenly recalled as contextual traits, and thus not be integrated in the focal mean. Alternatively, focal traits whose favorability was opposite to that of the contextual group may have appeared particularly salient, and thus may have received greater weight in the averaging process.

To tap subjects' memory for traits, a surprise-recall task was presented at the end of the experiment. Subjects were asked to list all the traits they could remember and place them in the appropriate group A or B. Indeed, there was impaired recall of those traits whose favorability would allow association with either group. The mean favorability of focal traits that were incorrectly recalled as contextual traits was virtually identical to the mean favorability of the contextual traits ( $M$ , with negative context = 27.37;  $M$ , with positive context = 80.14). That is, the favorabilities of 12.7% of the focal traits distorted mean estimates away from the contextual groups because these traits were not associated with that group in memory. When averaged, the traits that were correctly placed in the focal group showed even greater category accentuation than mean estimates. More negative



traits were recalled when the context was negative ( $M = 73.43$ ) than when it was positive ( $M = 37.41$ ). This result was surprising, because the assumption was that biased retrieval would explain but not surpass biased mean estimates.

Ideally, recall would have been studied at the time of mean estimation. Memory data collected at the conclusion of the experiment are suggestive, but may involve additional sources of error. At the end of the experiment, the categorization into groups A or B (and the affective connotation associated with them) served as the sole retrieval cue other than the semantic traces of the individual traits. In phase 1, subjects had learned that there was a high biserial correlation between trait favorability and group membership. This correlation was attenuated in phase 2 with the increase in variance and the emerging overlap in favorability between groups. Is it possible that subjects relied on a simple decision rule such as "All the nice people were in group A and all the nasty people were in group B?" Perhaps. One set of data, however, suggests that such a rule may only partially explain these findings. When asked to recall the lowest focal rating they had made, subjects' recalled ratings were more favorable ( $M = 12.73$ ) than their actual ratings ( $M = 4.46$ ), but recall of the highest ratings ( $M = 92.41$ ) was almost as positive as actual highest ratings ( $M = 96.27$ ). Thus, at least on the level of ranges of favorability, the two distributions were not remembered as mutually exclusive. There remains the possibility, however, that in hindsight the degree of perceived intergroup overlap, in terms of variances, was smaller than it was during the experiment.

Before exploring alternatives to the memory explanation, I will review further evidence for the ideas that: (a) category-accentuation effects are not specific to judgments about groups of people; and (b) contrast and accentuation effects arise at different stages of information processing. The first issue was addressed in an experiment which was similar to the one described above. Subjects were presented with the same set of trait adjectives, but instead of rating the likeability of group members, they rated the desirability of the presented words and then estimated the mean desirability for each list, A and B (Krueger & Rothbart, 1990, Experiment 2). Results were practically identical to the ones reported above, suggesting that the cognitive processes involved in category accentuation are not specific to judgments about groups of people.

To differentiate processes in category accentuation better, contrast effects were attributed to distorted ratings of individual stimuli, and accentuation effects were attributed to differences in judgmental weight given to stimuli in the averaging process. By this definition, contrast effects are only possible when stimuli are sufficiently ambiguous to be perceived or reproduced differently, depending on changes in the context. To address the second

issue, the independence of these two kinds of effects, several experiments were conducted in which individual stimuli were unambiguous and could not be distorted.

### Number Experiments

In one experiment, 96 three-digit numbers, classified into two mutually exclusive categories (A and B), were the exemplars (Krueger, 1991, Experiment 2). Numbers did not carry any meaning beyond their apparent value. That is, no mention was made that the numbers might involve characteristics of human groups. Subjects were informed that they participated in an experiment on intuitive averaging. Again, there were two phases. In phase 1, the focal category (range 147–164;  $M = 155.5$ ) was presented either with smaller contextual numbers (range 129–145;  $M = 137.5$ ) or with larger contextual numbers (range 165–182;  $M = 173.5$ ). Subjects typed in each number presented, and periodically estimated the cumulative mean for each category. The frequency of errors in number recognition was negligible and there were no contrast or accentuation effects. When the variance of numbers was increased in the focal category in phase 2 (range 129–182), however, extreme stimuli appeared again to be carrying greater weight in mean estimation, and thus intercategory differences were accentuated. When the contextual category comprised small numbers, subjects perceived an illusory increase of 2.02 points in the mean. When the contextual category comprised large numbers, there was an illusory decrease of 1.44 points.

The illusory perception of increased intercategory differences is but one manifestation of accentuation. In online category learning, true means need not necessarily be stable. For example, growing familiarity with a new social group may entail true changes in the mean characteristic of that group. If accentuation is operating in such a situation, shifts in estimated means should be greater when the group becomes more dissimilar to some comparison group than when it becomes more similar. Numbers served again as stimuli and phase 1 was comparable to phase 1 of the experiment described above (Krueger, Rothbart, & Sriram, 1989, Experiment 1). No contrast or accentuation effects were observed. In phase 2, the true means either increased or decreased by 6 points. Therefore, true intercategory differences were *enhanced* when focal means increased and contextual numbers were small, or when focal means decreased and contextual numbers were large. Conversely, true differences were *reduced* when focal means increased and contextual numbers were large, or when focal means decreased and contextual numbers were small. Reductions of mean differences (unlike enhancement) also resulted in substantial intercategory overlap of numbers.

The first goal of this experiment was to study accentuation effects in the



change of mean estimates. If difference-enhancing stimuli carry greater weight in mean estimation than difference-reducing stimuli, changes in estimated means should differ despite the equal amount of true change. The second goal was to explore the role of intercategory proximity. To do this, a condition was introduced in which the contextual category did not border the focal category but was removed by several hundred points. It was hypothesized that only proximal contextual categories affect judgments of focal means.

Estimates of category means were biased only when the two distributions were sufficiently close. Changes in mean estimates were indeed greater when true differences were enhanced ( $M = 9.00$ ) than when they were reduced ( $M = 6.20$ ). When the contextual category was distant, it did not matter whether intercategory differences became smaller or greater, and the average changes in estimated means fell between the conditions of enhancement and reduction ( $M = 7.16$ ). Apparently, both the greater weight of extreme, difference-enhancing stimuli and the diminished weight of difference-reducing stimuli contributed to the accentuation of change. All changes in estimated means were greater than the average change in the true means ( $M = 4.16$ ). This effect reflected superior recall for recent information. All numbers entered the calculation of the true means with equal weight, but because the average-modifying numbers were concentrated in the second half of the experiment, it was not surprising that they were more accessible for subjects when they estimated the cumulative means.

The third goal was to track changes throughout phase 2. Subjects gave estimates six times during this phase, after each set of four focal and four contextual numbers. For each of these measurement times a difference score, indicating perceived change, was computed. Estimates in phase 1 were averaged across subjects and then subtracted from averaged estimates in phase 2. The sign of the difference was ignored. Because the direction of true change turned out to be irrelevant, data were pooled across the conditions of upward and downward change. Figure 2.1 shows the average changes throughout phase 2 for the enhancement, reduction, and baseline conditions, and the true change.

True changes increased monotonically and their trajectory was slightly negatively accelerated. Comparisons between the shapes of the trajectories of true and empirical changes revealed no significant differences. That is, the temporal patterns of gradual modifications of mean estimates were highly accurate, even though overall change was overestimated. At least on the level of group statistics, exemplar-based category change supported the book-keeping model of stereotype change rather than the conversion model (Rothbart, 1981). The findings of this study were replicated in an experiment using trait stimuli (Krueger & Rothbart, 1990, Experiment 1).

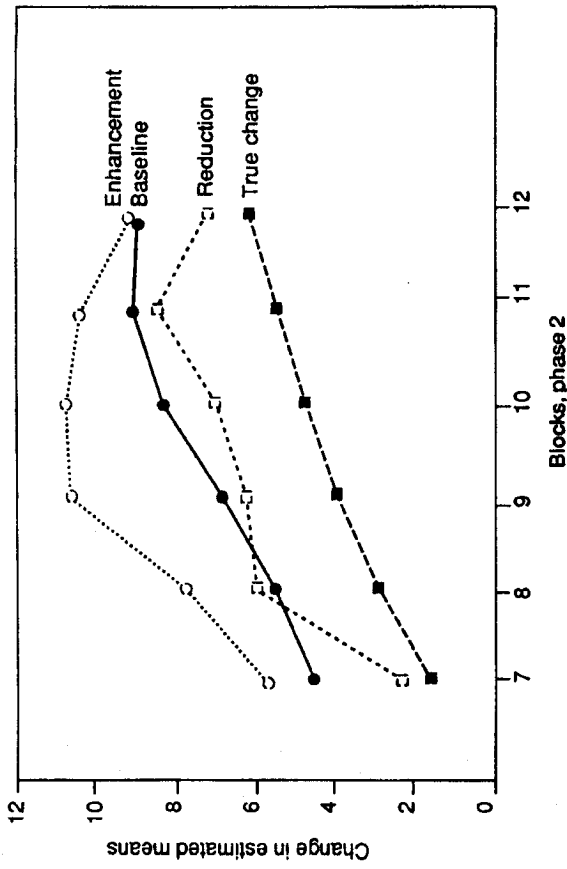


Figure 2.1 Difference scores (ratings in phase 2 minus the average of phase 1) for the enhancement condition, the baseline condition, the reduction condition, and true change

### The Role of Perception, Memory, and Belief Verification

Taken together, the reported trait and number experiments suggested that accentuation effects should be distinguished from contrast effects. Contrast effects referred to perceptual distortions of individual category exemplars, whereas accentuation effects referred to any additional spreading apart of estimated category means that could not be explained by accurate averaging of distorted individual ratings.

Three different kinds of processes may explain these accentuation effects. First, the distinctiveness principle suggests that extreme, outlying stimuli may be perceptually salient because of their uniqueness, and thus may be more available in memory (Tversky & Kahneman, 1973; von Restorff, 1932). As noted earlier, subjects may have selectively encoded or retrieved stimuli from the ends of the scale.

Second, category membership of stimuli located in the region of intercategory overlap may have been incorrectly recalled in a way that sharpened the boundaries (see Campbell, 1956). So far, the results presented from our paradigm are partially compatible with both processes, but they are not fully satisfactory.

A third possibility involves the notion that subjects formed a belief about intercategory distinctions in phase 1 and that they expected this belief to be

corroborated in phase 2. Recall that parameters of the stimulus distributions remained constant during phase 1, while only the number of stimuli increased. In phase 2, on the other hand, there was actual and unexpected change. Imagine the sequence of belief formation and change in a hypothetical setting. Assume an anthropologist encounters the Boringians and the Begoodians on a remote island. The two tribes share the same culture and language, and tribe membership can only be identified by some sartorial characteristic. As she gradually comes to know members of these two tribes, the anthropologist observes a second difference. The Begoodians are more affable than the Boringians. In fact, no Begoodian is less friendly than any Boringian. Consequently, friendliness becomes an important characteristic by which the tribespeople can be distinguished. As her acquaintance with the natives deepens, the anthropologist's experience may begin to filter through a stereotype. She may come to *expect* all Begoodians to be friendlier than Boringians and discard evidence to the contrary. She may not have discarded the same evidence if it had been present during her initial contact.

As a preliminary test of the two-stage process model of expectancy formation and accentuation, an experiment was conducted with numbers as stimuli (Krueger, Rothbart & Sriram, 1989, Experiment 2). The numbers presented were identical with the numbers used in the previous experiment. The only difference was that the focal and the contextual categories either overlapped from the start or they did not overlap. That is, there was no separation of a learning phase from a change phase. According to the distinctiveness hypothesis, accentuation effects should nevertheless occur, because of privileged accessibility of extreme numbers. Similarly, the minimization-of-overlap hypothesis predicts accentuation because of confusion or forgetting of numbers close to the boundary. In contrast, the hypothesis of belief formation and change predicts no accentuation effects when the distributional characteristics are constant across time. The results supported the hypothesis that stereotypic expectations about differences affect the integration of new, average-modifying information because in this experiment no accentuation effects were found.

### Automatic and Controlled Thought in Category Learning

It is difficult to evaluate to what extent expectations regarding intergroup differences operate via unintentional automatic mechanisms or via controlled and motivated thought. Tajfel (1969) pointed out that high correlations between categorical variables like group membership, and continuous variables like personal attributes (e.g. friendliness or height), are advantageous because they maximize predictive utilities. Believing that all Begoodians are friendlier than the Boringians, the imaginary anthropologist may be motivated to keep it that way and to "repress" contradictory

information. Alternatively, information that blurs perceived category boundaries may be automatically and unintentionally dropped. The available evidence tends to support the latter alternative.

To test whether categorization effects can be eliminated by inducing subjects to put a premium on perceptual accuracy, it might be necessary to reward unbiased responses. If subjects were encouraged, for example by the promise of monetary rewards, to judge the length of lines precisely or to rate the favorability of traits independently of other presented traits, one could determine the extent to which such judgments can be improved by controlled thought. Clarke and Campbell (1955) found no significant intergroup contrast when an award of \$5.00 was offered to the student who would predict test scores of Blacks and Whites most accurately. No such experiment has been conducted, however, in a categorization procedure not involving ingroup-outgroup distinctions. The dual-mean estimation task in the trait and number studies involved procedures with maximum demand on controlled and effortful cognition. Subjects' intent to provide accurate judgments became evident in informal postexperimental questioning. Many were spontaneously interested in the quality of their performance, and some expressed disappointment about the lack of immediate feedback about the accuracy of their estimates.

Another possibility of exploring the role of intended distortions involves comparisons between the learning of social and non-social categories. We had initially speculated that learning about social groups would strengthen accentuation effects. Potential motivations to maximize the predictive utility of categories might be activated when subjects learn socially meaningful information about groups of people, rather than *ad hoc* psychophysical categories. Several experiments demonstrated that this was not the case. When categorized numbers were presented as measures of body weight of two groups of athletes (marathoners and sprinters), distortions were not greater than when the numbers were meaningless. Even when an abstraction-based stereotype was provided (i.e. the experimenter explicitly informed subjects that sprinters should be expected to be heavier than marathoners) the bias was not greater than when exemplar-based information was presented alone (Krueger, Rothbart & Sriram, 1989, Experiment 3). Similarly, when numbers differentiated the favorability of two groups (they indicated intelligence scores of two groups of students), no increase in bias was found (Krueger, 1991, Experiment 1).

Davis-Stitt, Rothbart and Krueger (1991) hypothesized that subjects perceive especially large between-group differences when they are categorized as members of one of the two groups. Adopting a minimal-group paradigm, we asked subjects to fill out a questionnaire which ostensibly measured the personality dimension of "leveling" versus "sharpening." In fact, items were constructed in an off-hand manner and not intended to measure anything;

yet, they were meant to look like real personality-inventory items (e.g. "When I read a novel, I look at the last page first to see how it ends"). Subjects in the experimental conditions were then given false feedback and randomly categorized as either levelers or sharpeners. In the control conditions, subjects were not categorized. Then all participants performed the standard mean estimation task. They were led to believe that the focal and contextual numbers represented scores of levelers and sharpeners on some cognitive test. In all conditions, the basic category-accentuation effect was replicated, but the effect was not larger for subjects who had been assigned group membership. It should be noted that in this study between-group difference had no evaluative connotations. It is conceivable that ingroup status magnifies categorization effects when the ingroup members are characterized by more desirable attributes than outgroup members, and that such effects are minimized when ingroup members are characterized by less desirable attributes.

## CONCLUSIONS AND IMPLICATIONS

Next to ingroup favoritism and perceptions of group homogeneity, the overestimation of between-group differences was introduced as the third persistent phenomenon in stereotyping. A series of studies demonstrated a persistent tendency to exaggerate between-category distinctions. In line with the initial assumptions, it was shown that category-accentuations:

- (a) Were present when learning was strictly *exemplar-based*.
- (b) Affected the computed and estimated *central tendencies*.
- (c) Highlighted the need to conceptualize stereotypes in the *context* of multiple category learning.
- (d) Presupposed *no ingroup-outgroup distinctions*, or any other kind of social categorization.

Taken together, these studies advance the understanding of the perception of intergroup differences in two ways. First, perception-based contrast effects were distinguished from cognition-based accentuation effects. Second, the accentuation of true change and the perception of illusory change were demonstrated for the first time. Employing a trial-by-trial learning procedure, gradual shifts in the critical variable (central tendency) were tracked online. Dynamic methods have already been used in prototype learning (Busemeyer & Myung, 1988), and they hold promise for intergroup research to expand from the area of stereotype formation into the area of stereotype change. To conclude this chapter, a few comments are due concerning the

relevance of this phenomenon for other areas of social perception and conflict.

### Effects of Stereotypes on Person Perception

How do people form impressions about others? Naturally, directly observed behaviors have the greatest influence on judgments about personality. Someone who consistently appears late for class is seen as unpunctual and perhaps as generally unreliable. Do stereotypes, to the degree that they involve perceived intergroup differences, affect inferences about individuals? Locksley *et al.* (1980) expected that women who had behaved assertively (e.g. telling a seedy character in the street to leave them alone) would still be seen as less assertive than men who had exhibited the same behavior. However, both target persons were judged to be equally assertive. Locksley *et al.* (1980) explained this finding in terms of a judgmental base-rate fallacy. People's gender stereotypes, they argued, imply differences in estimated probabilities of certain behaviors. When concrete individuating information is available, these implicit base rates are ignored (see also Kahneman & Tversky, 1973).

Krueger and Rothbart (1988) replicated the base-rate fallacy with weak stereotypes and strong diagnostic individuating information. However, when the stereotype was strong, or when the individuating information was only moderately diagnostic of the criterion, stereotypes affected judgments of personality. The stronger the stereotypic gender differences, the greater was their effect on trait attribution. For example, men were seen as more aggressive than women, and this difference was accentuated if the men were construction workers and the women were homemakers. To the extent that gender differences are overestimated, the base-rate fallacy may be offset, and stereotypes may contribute to judgments about individuals.

### Stereotype Change and Conflict Reduction

It may not always betray unwarranted favoritism when ingroups are perceived to differ favorably from outgroups; but when it can be shown that differences along an evaluative dimension are overestimated, such perceptions can be particularly damaging. In the context of attitude perception, Dawes, Singer and Lemons (1972) described the following vicious circle:

To the degree to which this [overestimation of differences] is strong or prevalent, it will exacerbate the conflict between opposing groups; for if members of each group believe that statements of members of the other group represent more extreme attitudes than they in fact do, each group will believe it to be more difficult to compromise with the other than it in fact would be; further, the belief that members of the other group hold extreme attitudes

will lead to a lowering of esteem for that group, which in turn will strengthen the contrast effect. Thus, this effect is both destructive and self-perpetuating (p. 281).

Given this bleak analysis, how can stereotypes be improved? Two models of change have been proposed, both including a dated traditional version as well as a more promising (and more modest) recent formulation. First, the *contact hypothesis* (Amir, 1976) in its original form cannot be maintained (e.g. Hewstone & Brown, 1986). Contact *per se* is insufficient to improve intergroup attitudes. A recent 4-year study again clearly corroborated the enormous stability of ethnic, racial, and occupational stereotypes in a college population (Rothbart & John, in press). The experiments reported in this chapter, however, suggest that under special circumstances stereotypes can change significantly; namely, when they are exemplar-based and when all the new and relevant information suggests a shift of the group mean in the same direction. In that case, perceived means change more than the actual means, even when the direction of change reduces between-group differences.

Second, a simple form of the *recategorization hypothesis* may demand all categorization to be eliminated, and beliefs about social groups to be supplanted by global conceptions about humanity, or by highly individualized conceptions about specific persons. At the time of the Cold War, White (1957) suggested that the perception of a bimodal distribution of social attitudes in socialist and capitalist countries was illusory. Instead, he argued, people in both types of society shared a "modal philosophy" of common values of political freedom and economic justice. Acknowledging the commonalities of public opinion would take the edge out of the international conflict. As an alternative to replacing stereotypes by superordinate categories, such as humanity, Langer, Bashner, and Chanowitz (1985) proposed to decrease prejudice by increasing discrimination. According to this view, mindful attention paid to individuating characteristics diminishes the likelihood that stereotypic expectations will influence judgments about persons.

A problem of these methods of recategorization may be that they underestimate the automaticity of categorization and its benefits. Sandwiched between the superordinate category "humanity" and the subordinate category "individual," stereotypes may be the basic-level categories of social perception (Oakes & Turner, 1990). As is well known from the object-categorization literature, neither superordinate nor subordinate levels achieve the predictive power of basic-level categories.

Common-enemy theories acknowledge the importance of categorization. Modern versions of this ancient idea (Spartans and Athenians ignored their differences in view of the Persian menace) show that assimilation (reduction of intergroup differences) can be achieved through contrast. The rationale

is simple. An outgroup will be perceived as relatively similar to the ingroup when another, even more dissimilar outgroup becomes salient. Doise, Deschamps and Meyer (1978) showed that perceived ethnic differences between Swiss linguistic groups (German, French, Italian) were attenuated when a third outgroup was introduced that differed both linguistically and nationally (e.g. Germans from Germany or French from France). Wilder and Thompson (1988) elegantly illustrated this assimilation process in a mock-jury paradigm. When a moderately different outgroup was somewhat more similar to the ingroup than to an extremely different outgroup, its views were assimilated to the views of the ingroup. Otherwise, both outgroups were seen as similar to each other and very different from the ingroup. It may be hoped that conflicts with proximal and moderately different outgroup can be effectively curbed by the salience of other more distant and different groups. However, in a world of global alliances and confrontations, the juxtaposition of a third group may lead to Orwellian scenarios of conflict substitution instead of conflict resolution.

## ACKNOWLEDGMENTS

I wrote this chapter while I was a postdoctoral research fellow at the Max Planck Institute for Human Development and Education in Berlin. I thank Jutta Hundertmark for the preparation of Figure 2.1. Chuck Stangor and two anonymous reviewers helped shape this chapter with thoughtful suggestions.

## REFERENCES

- Allport, G.W. (1954). *On the Nature of Prejudice*. Reading, MA: Addison-Wesley.
- Amir, Y. (1976). The role of intergroup contact in change of prejudice and ethnic relations. In P. Katz (Ed.), *Toward the Elimination of Racism*. New York: Pergamon Press, pp. 245-290.
- Amster, H. (1964). Evaluative judgment in incidental learning. *Journal of Verbal Learning and Verbal Behavior*, 3, 466-473.
- Biernat, M., Manis, M., & Nelson, T.E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, 60, 485-499.
- Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3, 27-52.
- Brown, D. (1953). Stimulus-similarity and the anchoring of subjective scales. *American Journal of Psychology*, 66, 199-214.
- Bussemeyer, J.R., & Myung, I.J. (1988). A new method for investigating prototype learning. *Journal of Experimental Psychology*, 14, 3-11.
- Campbell, D.T. (1956). Enhancement of contrast as composite habit. *Journal of Abnormal and Social Psychology*, 53, 350-355.

- Campbell, D.T. (1967). Stereotypes and the perception of group differences. *American Psychologist*, 22, 817-829.
- Clarke, R.B., & Campbell, D.T. (1955). A demonstration of bias in estimates of Negro ability. *Journal of Abnormal and Social Psychology*, 51, 585-588.
- Davis-Stitt, C.L. (1989). Categories and continua: Effects of arbitrary boundaries on perceived similarity. Unpublished manuscript, University of Oregon.
- Davis-Stitt, C.L., Rothbart, M., & Krueger, J. (1991). The accentuation of category boundaries in the minimal-group paradigm. Unpublished data, 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.
- Diehl, M. (1990). The minimal group paradigm: theoretical explanations and empirical findings. In W. Stroebe & M. Hewstone (Eds), *European Review of Social Psychology*, Vol. 1. Chichester: Wiley, pp. 263-292.
- Doise, W., Deschamps, J.-C., & Meyer, G. (1978). The accentuation of intra-group similarities. In H. Tajfel (Ed.), *Differentiation Between Social Groups*. London: Academic Press.
- Eiser, J.R. (1971). Enhancement of contrast in the absolute judgment of attitude statements. *Journal of Personality and Social Psychology*, 17, 1-10.
- Freud, S. (1959, originally published in 1921). *Group Psychology and the Analysis of the Ego*. New York: W.W. Norton & Co.
- Goldberg, L.R. (1973). Some Normative Characteristics of the 1710 Trait-descriptive Adjectives. Unpublished manuscript, Oregon Research Institute, Eugene, OR.
- Helson, H. (1964). *Adaptation-level Theory*. New York: Harper & Row.
- Hewstone, M., & Brown, R.J. (1986). Contact is not enough: An intergroup perspective on the "contact hypothesis." In M. Hewstone & R.J. Brown (Eds), *Contact and Conflict in Intergroup Encounters*. Oxford: Blackwell, pp. 1-44.
- Howard, J., & Rothbart, M. (1980). Social categorization and memory for ingroup and outgroup behavior. *Journal of Personality and Social Psychology*, 38, 301-310.
- Judd, C.M., & Park, B. (1988). Outgroup homogeneity: judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology*, 54, 718-788.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Katz, D., & Braly, K.W. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, 28, 280-290.
- Kenrick, D.T., & Gutierrez, S.E. (1980). Contrast effects and judgments of physical attractiveness: when beauty becomes a social problem. *Journal of Personality and Social Psychology*, 38, 131-140.
- Kenrick, D.T., Gutierrez, S.E., & Goldberg, L. (1980). Influence of popular erotica on judgments of strangers and mates. *Journal of Experimental Social Psychology*, 25, 159-167.
- Klineberg, O. (1944). *Characteristics of the American Negro*. New York: Harper.
- Krueger, J. (1991). Accentuation effects and illusory change in exemplar-based category learning. *European Journal of Social Psychology*, 21, 37-48.
- Krueger, J., & Rothbart, M. (1988). Use of categorical and individuating information in making inferences about personality. *Journal of Personality and Social Psychology*, 55, 187-195.
- 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.
- Langer, E., Bashner, R., & Chanowitz, B. (1985). Decreasing prejudice by increasing discrimination. *Journal of Personality and Social Psychology*, 49, 113-120.
- Lewis, S.B. (1990). Ideals as prototypes. Unpublished doctoral dissertation, University of Oregon.
- Lilli, W., & Lehner, F. (1971). Stereotype Wahrnehmung: eine Weiterentwicklung der Theorie Tajfels. *Zeitschrift für Sozialpsychologie*, 2, 285-294.
- Linville, P.W., Fischer, G.W., & Salovey, P. (1989). Perceived distributions of the characteristics of ingroup and outgroup members: empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, 57, 165-188.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, 39, 821-831.
- Locksley, A., Ortiz, V., & Hepburn, C. (1980). Social categorization and discriminatory behavior: extinguishing the minimal intergroup discrimination effect. *Journal of Personality and Social Psychology*, 39, 773-783.
- McCauley, C., & Stitt, C.L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, 36, 929-940.
- McGarty, C., & Penny, R.E.C. (1988). Categorization, accentuation, and social judgement. *British Journal of Social Psychology*, 27, 147-157.
- Oakes, P.J., & Turner, J.C. (1990). Is limited information processing capacity the cause of social stereotyping? In W. Stroebe & M. Hewstone (Eds), *European Review of Social Psychology*, Vol. 1. Chichester: Wiley, pp. 112-135.
- Park, B., & Hastie, R. (1987). Perception of variability in category development: instance- versus abstraction-based stereotypes. *Journal of Personality and Social Psychology*, 53, 621-635.
- Park, B., & Rothbart, M. (1982). Perception of outgroup homogeneity and levels of social categorization: memory for the subordinate attributes of ingroup and outgroup members. *Journal of Personality and Social Psychology*, 42, 1051-1068.
- Petzold, P. (1990). Einfluss von Klassifizierungen auf die Beurteilung von Eigenschaften. *Zeitschrift für Psychologie*, 198, 59-68.
- Rogers, S. (1941). The anchoring of absolute judgments. *Archives of Psychology*, 261.
- Rokeach, M. (1960). *The Open and the Closed Mind*. New York: Basic Books.
- Rothbart, M. (1981). Memory processes and social beliefs. In D.L. Hamilton (Ed.), *Cognitive Processes in Stereotyping and Intergroup Behavior*. Hillsdale, NJ: Erlbaum, pp. 145-182.
- Rothbart, M., & Hallmark, W. (1988). Ingroup-outgroup differences in the perceived efficacy of coercion and conciliation in resolving social conflict. *Journal of Personality and Social Psychology*, 55, 248-257.
- Rothbart, M., & John, O.P. (in press). Intergroup relations and stereotype change: a social-cognitive analysis of some longitudinal findings. In P.M. Sniderman & P.E. Tetlock (Eds), *Prejudice, Politics, and Race in America*. Stanford: Stanford University Press.
- Secord, P.F., Bevan, W., & Katz, B. (1956). The negro stereotype and perceptual accentuation. *Journal of Abnormal and Social Psychology*, 53, 78-83.
- Sherif, M., Taub, D., & Hovland, C.I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, 2, 150-155.
- Simon, B., & Brown, R. (1987). Perceived intragroup homogeneity in minority-majority contexts. *Journal of Personality and Social Psychology*, 53, 703-711.
- Simon, B., & Pettigrew, T.F. (1990). Social identity and perceived group

- homogeneity: evidence for the ingroup homogeneity effect. *European Journal of Social Psychology*, 20, 269-286.
- Sumner, W.G. (1906). *Folkways*. Boston: Ginn.
- Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, 25, 79-97.
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96-102.
- Tajfel, H., & Wilkes, A.L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, 54, 101-114.
- Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Von Restorff, H. (1932). Über die Wirkung von Bereichsbildungen im Spurenfeld [On the effects of category formation in the field of memory traces]. In W. Koehler & H. von Restorff (Eds), *Analyse von Vorgängen in Spurenfeld: I. Psychologische Forschung*, 18, pp. 229-342.
- White, R.K. (1957). The cold war and the modal philosophy. *Conflict Resolution*, 11, 43-50.
- Wilder, D.A., & Allen, V.L. (1978). Group membership and preference for information about others. *Personality and Social Psychological Bulletin*, 4, 106-110.
- Wildel, D.A., & Thompson, J.E. (1988). Assimilation and contrast effects in the judgments of groups. *Journal of Personality and Social Psychology*, 54, 62-73.
- Winter, D.G. (1987). Enhancement of an enemy's power motivation as a dynamic of conflict escalation. *Journal of Personality and Social Psychology*, 52, 41-46.