JOURNAL OF
Personality
AND
Social Psychology

Volume 36 September 1978 Number 9

An Individual and Quantitative Measure of Stereotypes

Clark McCauley
Bryn Mawr College

Christopher L. Stitt
Temple University

A quantitative and individual measure of stereotyping is proposed, based on *defining stereotypes as probabilistic predictions that distinguish the stereotyped* group from others. Data indicate that the proposed measure, though related to the familiar Katz and Braly checklist, is a substantially new measure of stereotyping rather than simply a quantitative version of the checklist. It is argued that the theoretical and empirical value of the proposed measure is justification for abandoning the Katz and Braly measure. Theoretically, the new measure relates stereotype research to attribution theory as part of a Bayesian approach to the psychology of prediction. Empirically, the new measure opens interesting questions about stereotypes, especially about stereotype validity and the "kernel of truth" hypothesis. A study using the new measure reveals that diverse groups of subjects have some similar stereotypes of black Americans, that these stereotypes are relatively accurate, and that contrary to the "kernel of truth" hypothesis, these stereotypes are seldom exaggerated.

Empirical study of stereotypes began with Katz and Braly (1933) and goes on today with the same measure they first used with Princeton students. Subjects are given a list of traits and asked to check the five traits "most typical" of a particular group. A stereotype is said to exist to the extent subjects agree on the choice of traits. This stereotype is properly a social stereotype because it is a group measure, defined by agreement across subjects.

The authors wish to acknowledge the participation of Gerry Ritrovato and Terryll Schasse in carrying out the study of stereotypes of black Americans (Study 3 in this article).

Requests for reprints should be sent to Clark McCauley, Department of Psychology, Bryn Mawr College, Bryn Mawr, Pennsylvania 19010.

Brigham (1971a), reviewing the stereotype literature since 1933, concluded that little has been discovered about how stereotypes are learned, changed, or affect behavior. According to Brigham, these questions go unanswered (and almost unasked) because of the lack of an individual and quantitative measure of stereotyping. Stereotypes exist in individuals, but, as noted above, the checklist can measure only a kind of group-average stereotype.

A Stereotype Differential measure suggested by Gardner, Kirby, Gorospe, and Villamin (1972) does provide a more quantitative version of the checklist, by requiring subjects to rate a particular group on a number of bipolar trait scales. But Gardner et al. still depend on group consensus to define the stereotype: They use the t statistic to find those (stereotype) scales on which the group mean rating is most reliably different from bipolar neutrality. Thus, the Stereotype Differential is like the checklist in measuring a social rather than an individual stereotype. The two are also alike in being limited to trait adjectives as the stereotype characteristics of interest.

In an effort to provide a better stereotype measure, Brigham (1971a) suggested that subjects be asked to judge a within-group probability, the percentage of a stereotyped group thought to have any trait or characteristic of interest. Brigham has used the percentage measure in a number of recent studies (Brigham, 1969, 1971b, 1973, 1974), and several interesting results have emerged. Contrary to the assumption of many investigators, the trait adjectives chosen as "most typical" with the Katz and Braly checklist are far from exceptionless generalizations. "In many cases, subjects saw *all five* of the 'most typical' traits as being characteristic of *less than half* of the members of the ethnic group" (Brigham, 1971a, p. 30). Furthermore, the traits for which the mean percentage is highest can turn out to have the least relation to attitude toward the target ethnic group (Brigham, 1971a). These results are somewhat surprising and raise the possibility that Brigham's percentage measure may be getting at something different than the checklist method does, rather than being simply a quantitative refinement of the checklist method.

The possibility that the percentage measure is not a stereotype measure is barely suggested, however, by the surprisingness of some of the data resulting from use of the measure. In fact, only a better measure can make clear where the within-group percentage measure is deficient. The discussion thus far has been an introduction for the new measure of stereotyping that is proposed in this article. The proposed measure emerged unexpectedly from data gathered in studies of the psychology of prediction. Once recognized, the new measure promises to be the quantitative and individual measure of stereotyping that Brigham was looking for but did not quite find.

Bayes' Rule and Human Intuition

Research in the psychology of prediction has been largely centered around Bayes' rule, which is a formula for revising predictions in the light of data. Given any two events, A and B, Bayes' rule states that $p(B/A) = p(B) \cdot p(A/B)/p(A) = p(B) \cdot LR$. In words, the probability of B, given that A has occurred, is the base-rate probability of B *times* the probability of A given occurrence of B *divided by* the probability of event A. Thus the probability of B, given A, differs from the a priori or base-rate probability of B by the factor LR, the likelihood ratio $p(A/B)/p(A)$. LR is often called a "diagnostic ratio" because it is a measure of the degree to which occurrence of A revises the probability of B. When LR is 1.0, the occurrence of A says nothing about the probability of B, that is, A has no diagnostic value. Similarly and symmetrically, $p(B/A)/p(B)$ is also a diagnostic ratio, a measure of the revision in $p(A)$ by the occurrence of B. In fact, the diagnostic value of Event B for the probability of A is the same as the diagnostic value of Event A for the probability of B: $p(A/B)/p(A) = p(B/A)/p(B) = LR$.

Bayes' rule is the rational, prescriptive, or normative theory for revising prediction according to data. The extent to which Bayes' rule is also a descriptive theory of human prediction behavior has been an issue of much research. In a recent contribution to this literature, Kahneman and Tversky (1973) hypothesized that people violate Bayes' rule by using a representativeness heuristic for category judgments. Their contention is that people predict the category most representative of (similar to) an instance and thereby err in ignoring the reliability of the instance and, especially, the prior probability of the category. In several different kinds of studies, Kahneman and Tversky found evidence that even mathematically sophisticated persons follow the representativeness heuristic rather than Bayes' rule. One of these studies depended on using occupational stereotypes, and it is here that our own studies of stereotyping began.

The relevant study of Kahneman and Tversky did not require the usual kind of stereotype predictions, in which personality prediction is made on the basis of information about national or ethnic group membership. Instead, subjects were given a personality description of "Tom W." and asked to predict what field of graduate study Tom W. was in. There were three groups in the study. One group (base-rate group) estimated the proportion of students in nine different fields of graduate study. Other subjects (similarity group) ranked the graduate fields according to the similarity of Tom W. to the typical student in each field. Still other subjects (prediction group) were given the description of Tom W. and asked to rank the graduate fields according to the likelihood that Tom W. was in each field. In studies of this type, Kahneman and Tversky found that the prediction group's ordering was always highly correlated with the similarity group's ordering but only inconsistently and weakly correlated with the ratings of the base-rate group.

Failure to attend to base-rate information violates Bayes' rule, which requires that both base rate and representativeness be taken into account. In the study just described, the probability that a given personality is in Graduate Field_i is equal to the prior probability of Graduate Field_i *times* the probability of the given personality within Graduate Field_i *divided by* the prior probability of the given personality [$p(B/A) = p(B) \cdot p(A/B)/p(A)$]. The prior probability of the given personality is the same no matter what graduate field is being considered, so, across graduate fields, the category predictions should vary directly with prior probability of graduate field (base rate) and with probability of the given personality within the graduate field (similarity or representativeness). According to Bayes' rule, then, both base rate and similarity ratings should correlate with the stereotype predictions. Kahneman and Tversky showed that their subjects ignored the base rate.

Study 1: Bayes' Rule and Stereotype Predictions [1]

As noted above, the Kahneman and Tversky (1973) study depended on occupational stereotypes, but the prediction of interest was from personality to category (graduate field). We began with a study designed to test whether the more usual stereotype prediction from category information to personality would show the same failure to take account of base rate that Kahneman and Tversky had found. The new study used a national stereotype (Germans) rather than occupational stereotypes and asked subjects for direct estimation of the Bayesian probabilities rather than for rankings of these probabilities.

Method

Subjects

The subjects of the study were 69 Harcum Junior College females. These subjects were not mathematically sophisticated: Few had had calculus, and none had had statistics.

Materials

For this study we chose four traits from the social stereotype of Germans, as measured by Karlins, Coffman, and Walters (1969): *efficient, extremely nationalistic, industrious,* and *scientifically minded.* We listed these four with five other traits, not part of the German social stereotype, in alphabetical order. For each trait (e.g., *efficient*), we asked each subject to estimate three probabilities of Bayes' rule: p(efficient/German) was the "percent of Germans who are efficient"; p(efficient) was "the percent of all the world's people who are efficient"; and p(German/efficient) was "the percent of efficient people who are German." The fourth probability, p(German), was "the percent of the world's people who are German." Each kind of probability was assessed on a different page of a four-page questionnaire.

Procedure

The order of the pages of the questionnaire was varied for different subjects such that one third estimated p(trait/German) first, one third estimated p(trait) first, and one third estimated p(German/trait) first. All subjects estimated p(German) last.

The study was introduced to subjects as a study of quantitative prediction. Subjects were told that they would be asked questions they could not answer exactly, such as "What percent of American cars are Chevrolets?" Subjects were encouraged to guess even though unsure, just as they might guess that the percentage of Chevrolets is somewhere between 10% and 50%.

Results

Kahneman and Tversky (1973) asked each of their subjects only one kind of question. In order to make our results comparable to theirs, we first analyzed only the data from

[1] See McCauley, Stitt, and Greenberg (Note 1).

Table 1
Mean Probabilities in Percentages and Mean Diagnostic Ratios

Description	Judged p (trait)	Judged p (German/ trait)	Judged p (trait/ German)	Calculated p (trait/ German)	Diagnostic ratio
Efficient[a]	49.8	22.5	63.4	62.6	1.27
Extremely nationalistic[a]	35.4	23.6	56.3	46.7	1.59
Ignorant	34.0	11.9	29.2	22.6	.66
Impulsive	51.7	16.9	41.1	48.8	.79
Industrious[a]	59.8	30.4	68.2	101.6	1.14
Pleasure-loving	82.2	23.5	72.8	107.9	.89
Scientifically minded[a]	32.6	25.0	43.1	45.5	1.32
Superstitious	42.1	11.4	30.4	26.8	.72
Tradition-loving	62.4	22.2	57.2	77.4	.91

[a] Traits from German stereotype.

the first page of each questionnaire. These are "pure" data in the sense of being unaffected by any possible order effects in answering more than one kind of question. The first three columns of Table 1 present these pure data, the mean judged p(trait), p(German/trait), and p(trait/German) for each of the nine personality traits.

Product-moment correlations were computed between columns 2 and 3 and between columns 1 and 3. The correlation between p(German/trait) and p(trait/German) is .83 ($p < .05$). This result confirms the high correlation typically found by Kahneman and Tversky between their similarity and prediction questions. The correlation between p(trait) and p(trait/German) is .72 ($p < .05$). This result—a significant correlation between base rate and prediction—contradicts the typical result found by Kahneman and Tversky.

Kahneman and Tversky suggest that occasional correlations between base rate and prediction will occur when the most representative outcomes happen also to be relatively frequent (i.e., by accident of correlation between base rate and similarity judgments). In the present study, however, the correlation between p(trait) and p(German/trait)—first two columns of Table 1—is not significant ($r = .36$) and so cannot explain the correlation between base rate and prediction.

In the present study, unlike that of Kahneman and Tversky, a direct test of the subjects' use of Bayes' rule was possible. The fourth column of Table 1 presents the calculated Bayesian prediction, $p(\text{trait}) \cdot p(\text{German/trait})/p(\text{German})$, for each of the nine personality descriptions. These calculations were performed using the group means of columns 1 and 2 together with the mean p(German) calculated across all subjects (17.9%). A product-moment correlation between the directly judged prediction (column 3) and the calculated prediction (column 4) yielded $r = .91$ ($p < .05$). This result again indicates that the observed predictions followed closely the Bayesian logic of statistical prediction. Our subjects were making some absolute errors in judgment, however, as the calculated predictions in column 4 did twice exceed 100%.

As noted, the previous analyses were performed using group means. However, since each subject estimated all four probabilities of the Bayesian formula, it was also possible to test individuals' use of Bayes' rule. The four correlations just examined at the level of group means were calculated for each subject: her judged p(German/trait) correlated across nine traits with her judged p(trait/German); her judged p(trait) correlated with her judged p(trait/German); her judged p(German/trait) correlated with her judged p(trait); and her judged p(trait/German) correlated with the Bayesian p(trait/German) calculated from her other judgments. Using z transformations to average across the 69 subjects, the four mean correlations were, respectively: $r = .74$ ($p < .05$); $r = .68$ ($p < .05$); $r = .45$ (*ns*); and $r = .73$ ($p < .05$). These correlations confirm the group-mean correlations in showing both representativeness and base rate significantly correlated with the stereotype prediction but not significantly correlated with one another. The correlation between judged and calculated stereotype predictions ($r = .73$) is also significant at the individual level, though the size of this correlation is undoubtedly attenuated by the multiplication of error in the individual judgments multiplied for the Bayesian calculation.

Discussion

The surprising result of this study was that trait predictions about Germans, p(trait/German), covaried across traits with both base rate, p(trait), and representativeness, p(German/trait), as predicted by Bayes' rule. More abstractly, $p(B/A)$ covaried with $p(B) \cdot p(A/B)/p(A)$, where B was the set of persons fitting a trait description, A was the set of Germans, and the universe was all the world's people.

Two additional unpublished studies by McCauley, Stitt, and Greenberg have confirmed that people do take account of base rates in making stereotype predictions from group membership information. Both studies used the same procedure as the study described, except that the target group for both was Americans instead of Germans. The trait list used contained four traits taken from the American stereotype (Karlins et al., 1969) instead of from the German stereotype. Subjects in one study were 72 Montgomery County Junior College students (male and female); subjects in the other study were 48 Bryn Mawr undergraduates (female). The same kind of correlational analyses that showed German trait predictions consistent with Bayes' rule in the first study also showed American trait predictions to be consistent with Bayes' rule in these two additional studies.

It is not clear why the subjects in our studies were able to make predictions in accord with Bayes' rule, whereas subjects in the Kahneman and Tversky (1973) studies did not. Perhaps stereotype predictions from category to trait are generally easier than predictions from trait to category. Or perhaps direct estimation of probabilities in percentages is easier than the rank ordering of probabilities required of Kahneman and Tversky's subjects. What is clear is that college students, at least, are capable of producing meaningful estimates of the Bayesian probabilities involved in stereotype predictions. Our data cannot be said to show that people are typically Bayesian; we have examined one variation of only one of the several kinds of studies in which Kahneman and Tversky showed systematic departures from normative theory. Rather than join the debate as to when people are and are not Bayesian in their predictions, we turned instead to a new measure of stereotyping suggested by the data of Table 1.

The Diagnostic Ratio as a Stereotype Measure

Notice that column 3 of Table 1 contains the kind of percentages advanced by Brigham (1969, 1971) as a stereotype measure: the percentages of Germans having each of the nine listed traits, including four traits from the German social stereotype. The percentages in column 3 are like those reported by Brigham in that they are far from exceptionless generalizations: the highest percentage is only 72.8. And the traits from the German stereotype are not always the ones seen as most probable. *Pleasure-loving*, for instance, has a mean judged probability that is higher than any of the four stereotype traits. As with Brigham's data, then, these data do not make it easy to accept p(trait/German) as a stereotype measure.

In the fifth column of Table 1, however, is a measure that does bring out the German stereotype traits. Column 5 contains diagnostic ratios obtained by dividing mean judged p(trait/German) by mean judged p(trait) for each trait adjective. Column 5 shows that the stereotype traits differ from the nonstereotype traits in having mean diagnostic ratios greater than 1.0, that is, in being seen as *more probable for Germans than for the world in general*. It appears that the diagnostic ratio does better than Brigham's within-group percentage measure in distinguishing known stereotype traits.

Once the diagnostic ratio is examined as a stereotype measure, an understanding of its success in distinguishing stereotype traits begins to emerge. The stereotype traits used in the prediction study were those that the Karlins et al. (1969) subjects had most often checked as "typical" of Germans. The data of Table 1 indicate that subjects do not always interpret the word "typical" to mean "most likely"—as many psychologists, including Brigham, have assumed. Rather than picking the traits most likely for Germans, the Karlins et al. subjects seem to have picked the traits most "characteristic" or "distinctive" of Germans. A stereotype trait may be seen as characterizing a small percentage of Germans, far less than 100%, and may be seen as absolutely less probable in Germans than many other traits; it can still be a stereotype trait, can still be "typical" of Germans if it is seen as relatively more probable in Germans than in others.

The data of Study 1 thus led us to propose that the stereotype of a group is composed of all the predictions that distinguish the stereotyped group from others. In the Bayesian terms already introduced, a stereotype trait is any trait with a diagnostic ratio substantially different from 1.0.

Study 2: The Diagnostic Ratio in Relation to the Checklist Measure

The data presented in Table 1 strongly suggested that the diagnostic ratio provides an individual and quantitative measure of the same stereotypes that were obtained with the checklist. The evidence was only suggestive, however, because Study 1 was not designed to investigate the interrelation of possible stereotype measures. Study 1 did not include a checklist measure of stereotyping, and the discussion thus far has depended on comparing probability judgments from our subjects with checklist data from subjects studied years ago by Karlins et al. (1969). To determine the relation of the checklist measure, the diagnostic ratio measure, and Brigham's measure, all these measures had to be directly compared within a single group of subjects. Our expectation was that traits judged to be most typical of a group would be the traits with the highest diagnostic ratios, but that typical traits would be little related to Brigham's within-group percentage measure. The data of Study 1 seemed to say that people interpret "typical" to mean "distinctive" rather than just "likely."

Method

Subjects

The subjects of Study 2 were 33 female undergraduate psychology students at Douglas College and 19 female undergraduate students at Rosemont College. The two colleges are quite different: Douglas is relatively large and is state supported, whereas Rosemont is a small private (Catholic) school. Because students at the two schools differ in ways that might affect their stereotypes, data from the two schools were analyzed separately.

Procedure

Each subject ranked nine traits according to how typical they were of Germans and, for each trait, estimated the percentage of Germans with the trait (Brigham's measure) and the percentage of all the world's people with the trait (base rate). For half the subjects, the ranking was first and the probability estimates second, and for the other subjects, the order was reversed.

Materials

The selection of the nine traits for the trait list requires some justification. Study 1 focused on the interrelations of probability estimates, and the content of the trait list was important only insofar as it was necessary to ensure that the social stereotype traits for the target group were included in the trait list. Since the Douglas and Rosemont studies were designed to measure stereotypes, the content of the trait list is much more important. The content of a stereotype cannot be fairly assessed if the trait list omits important elements of stereotype description. Ideally, then, the trait list should include all 84 of the traits on the original Katz and Braly (1933) list.

Unfortunately, it is not practical to ask subjects for probability estimates for all 84 of the traits on the Katz and Braly list. In fact, it would be best if the number of traits on the list did not exceed nine, the number already used successfully. Jones and Ashmore (1973) have recently applied multidimensional scaling procedures to identify the underlying structure of stereotyped description, and it is possible to select eight adjectives to represent the dimensions of trait description that they uncovered: ***tradition-loving, artistic, superstitious, ignorant, industrious, competitive, materialistic,*** and ***nationalistic.*** The trait list included one additional adjective, ***pleasure-loving,*** because it was earlier useful in distinguishing between Brigham's (1969, 1971) within-class probability and the diagnostic ratio.

Results

The data from both Douglas and Rosemont colleges indicated that the ordering of questions (whether percentages or typicalness rankings were answered first) made no difference. Only two subjects, both from Douglas, professed to see no difference between *Germans* and *all the world's people* on all nine of the listed traits. These two subjects were retained in the analysis.

The major analysis was to calculate, for each subject, the correlation between typicalness ranking and diagnostic ratio, and between typicalness ranking and Brigham's within-group probability. Using z scores to average correlations across subjects, these two correlations were, respectively, −.49 and −.79 for Rosemont subjects and, respectively, −.66 and −.67 for Douglas subjects. (The correlations are negative because the more typical traits received the lower ranks.) All four correlations are significant ($p < .05$), but clearly they do not support our expectation that the diagnostic ratio is more related to typicalness than Brigham's within-group probability. For 16 of 19 Rosemont subjects and for 20 of 33 Douglas subjects, Brigham's measure correlated higher with typicalness than the diagnostic ratio did.

Discussion

Study 2 shows that the stereotype of Germans assessed by the checklist measure is related to but far from the same as the stereotype of Germans assessed by the diagnostic ratio measure. Our subjects evidently understood "typical" to mean both "distinctive" (ratio measure) and "likely" (Brigham's measure), rather than only "distinctive," as we had expected. Our expectation came from the neat fit between group diagnostic ratio data in Study 1 and group checklist data from Karlins et al. (1969). The individual data of Study 2 now make clear that the relation between the old measure and the new is not good enough to treat the diagnostic ratio as simply a quantitative form of the checklist measure. Although substantially related to the checklist measure, the ratio measure is also substantially different and must stand or fall independently as a useful measure of stereotyping.

After so many studies employing the Katz and Braly (1933) measure, it is difficult to avoid the feeling that a group's stereotype is whatever the checklist says it is and that any new measure must be consistent with the checklist if it is to be accepted as a stereotype measure. Study 2 indicates that the diagnostic ratio taps something rather different from what the checklist gets at. Nevertheless, we believe that stereotypes are best understood as predictions that distinguish the stereotyped group from others and that stereotypes are best measured as diagnostic ratios. The justification for abandoning the Katz and Braly measure is the empirical power and the theoretical integration offered by the new measure.

Empirical Value of the Diagnostic Ratio

The diagnostic ratio measure provides a quantitative and individual measure of stereotyping that should enormously facilitate the understanding of stereotypes and their effects. The checklist measure, a group measure of "social" stereotypes, has been very cumbersome to use. Consider studying some disease only from data of the incidence of that disease in different populations. Research might proceed, but only very awkwardly compared to what might be done if the disease process could be studied in individuals. The absence of an individual measure of stereotypes has meant that stereotypes have been studied only from essentially epidemiological data, and the result has been a dead end for stereotype research. We need to know who has how much of what stereotypes, how they were acquired, and what does or can change them. These questions can be investigated at the individual level by means of the diagnostic ratio measure.

The diagnostic ratio also opens the possibility of measuring negatively distinctive stereotypes. We have said that a group stereotype consists of predictions that distinguish that group from others, that is, predictions having diagnostic ratios different from 1.0.

Thus, a diagnostic ratio less than 1.0, as well as one greater than 1.0, can define a stereotype prediction. For instance, a trait with a diagnostic ratio of .5 is as strong as a stereotype as a trait with a diagnostic ratio of 2.0. In Table 1, *superstitious* (diagnostic ratio of .72) was about as negatively distinctive as *scientifically minded* (diagnostic ratio of 1.52) was positively distinctive. In other words, being less superstitious than other people was as much part of the German stereotype as being more scientifically minded than others.

Negatively distinctive stereotypings such as *superstitious* cannot be directly assessed with the checklist measure. As noted in Study 2, typical traits for a group appear to be determined on the basis of both high likelihood and positive distinctiveness. There may be, therefore, another side to stereotyping that has never been discovered with the checklist measure. Of course, the same stereotype that produces a diagnostic ratio less than 1.0 for *superstitious* might produce a diagnostic ratio greater than 1.0 for *hard-headed,* and perhaps *superstitious* as a negatively distinctive prediction is only the redundant reflection of *scientifically minded* as a positively distinctive prediction. Still, the difficulty of finding psychological opposites in any language is notorious, and some negatively distinctive stereotypings may be uncovered that cannot be alternatively assessed as positively distinctive stereotyping.

Study 3: Testing Stereotype Validity [2]

One particularly interesting question concerns stereotype validity and the "kernel of truth" hypothesis. LeVine and Campbell (1972) point out that stereotypes may not be totally invalid, but that perceptual processes are likely to exaggerate real differences between groups. The new measure makes possible direct testing of at least some stereotype predictions. All that is required is to ask people to estimate the probabilities of the diagnostic ratio for attributes with known probabilities in different groups.

Method

Subjects

We tested a number of groups, disparate in age and socioeconomic status, and analyzed the data separately for each group. We sought this variety for two reasons. First, we wanted to show that the diagnostic ratio measure is quite generally comprehensible and practical for different kinds of people. And second, we wanted to further the generalizability of our substantive results concerning the validity of stereotypes.

The subjects were 10 commercial-course students in a typing class at Roxborough High School in Philadelphia (all female; average age, 17 years); 17 Bryn Mawr, Haverford, and LaSalle College students (11 female, 6 male; average age, 20 years); 12 union members attending a leadership conference (2 female, 10 male; average age, 34 years); 10 members of the choir at Holy Angels Church in Philadelphia (9 female, 1 male; average age, 28 years); 13 Master of Social Work (MSW) students at the Bryn Mawr School of Social Work (10 female, 3 male; average age, 27 years); and 13 Bachelor of Science and MSW caseworkers employed by county welfare agencies in Pennsylvania (3 female, 5 male; average age, 32 years). There were four blacks among our subjects.

Materials

The questionnaire consisted of seven two-part questions. The first question was, "What percent of adult Americans have completed high school? What percent of adult American blacks have completed high school?" Six more questions asked, "What percent of (black) children born in America are illegitimate?"; "What percent of (black) Americans are *victims* of violent crimes?"; "What percent of (black) American families are on welfare?"; "What percent of (black) American families have four or more children under the age of 18 at home?"; and "What percent of (black) American families have a female head of family?"

Procedure

The high school class, the union class, the church choir, and the MSW class were approached as groups; the college students and the caseworkers were approached individually. Subjects were informed that the purpose of the study was to obtain subjective estimates of social statistics involved in important American social problems. It was emphasized that the answers to the questions were matters of verifiable fact and that the correct answers would be revealed after the questionnaire was completed. Subjects were asked to give their best estimates of the percentages asked on the questionnaire, just as they might guess at the percentage of American cars that are Chevrolets.

Results and Discussion

The diagnostic ratio measure of stereotyping was easily understood and completed. Although our subjects varied considerably in

[2] See McCauley and Stitt (Note 2).

Table 2
Criterion and Mean Diagnostic Ratios
(Percentage of Black Americans/Percentage of All Americans)

Characteristic	Criterion	High school (n = 10)	College (n = 17)	Union (n = 12)	Choir (n = 10)	MSW students (n = 13)	Caseworkers (n = 13)
1. Completed high school[a]	.65	**.68**	**.73***	**.67**	**.68**	**.60**	**.74**
2. Illegitimate	3.1	**1.8***	**1.7***	**2.1***	**1.9***	**2.3**	**2.0**
3. Unemployed last month[b]	1.9	1.9	**1.6***	**1.8**	**2.6**	**2.3**	**1.9**
4. Victims of crimes	1.5	.83	**1.8**	**2.0**	1.5	2.3	**1.9**
5. Welfare	4.6	**2.3***	**1.9***	**1.6***	**1.8***	**1.4***	**2.0***
6. Four or more children	1.9	**1.6**	**1.4***	1.6	**1.3***	**1.3**	1.4
7. Female head of family	2.8	**1.7***	**1.9***	1.7	**1.5***	**1.7**	**2.0**

Note. Bold face means are significantly different from 1.0 ($p < .05$, two-tailed, $df = n - 1$).
[a] See text for complete specification of these seven characteristics.
[b] Criterion ratio of 1.7 for Master of Social Work (MSW) students and caseworkers (data collected in a different month with different unemployment rates).
* Mean diagnostic ratio is significantly different from 1.0 ($p < .05$, two-tailed, $df = n - 1$) and significantly different from appropriate criterion ratio ($p < .05$, two-tailed, $df = n - 1$).

age and socioeconomic status, all were able to make estimates, in percentages, of the probabilities of the diagnostic ratio.

For each subject on each question, a diagnostic ratio was computed as follows (for example): percentage of black Americans completing high school/percentage of all Americans completing high school. Individual diagnostic ratios were then averaged across subjects within each group to obtain a mean diagnostic ratio for each group on each question. These mean diagnostic ratios are shown in Table 2.

The criterion diagnostic ratios appear in the first column of Table 2. These ratios were calculated from data obtained from *Statistical Abstract of the United States: 1975* (U.S. Bureau of the Census) and from U.S. Department of Labor reports (for last month's unemployment).

Of the 42 mean diagnostic ratios calculated, 37 (bold face in Table 2) are significantly different from 1.0 ($p < .05$, two-tailed). Our six groups of subjects thus show strong stereotyping of black Americans on the seven characteristics investigated. (Only two or three significant differences would be expected by chance in 42 t tests.) In all 37 cases of significant stereotyping, the mean diagnostic ratios differ from 1.0 in the same direction as their appropriate criterion ratios. Stereotyping on our seven characteristics is clearly veridical in terms of the direction of the difference perceived between black Americans and all Americans.

Contrary to the "kernel of truth" hypothesis (LeVine & Campbell, 1972), the diagnostic ratios do not appear to be consistently more extreme than the criterion ratios. Of the 37 mean diagnostic ratios indicating significant stereotyping, only 17 (starred in Table 2) are significantly different from their appropriate criterion ratios ($p < .05$, two-tailed, $df = n - 1$). All 17 of these diagnostic ratios are less extreme than their appropriate criterion ratios. There is not one instance of significant stereotyping where the stereotype ratio is more extreme than the criterion ratio.

It may be objected that our subjects did not reveal the full extremity of their stereotype, out of some kind of evaluation apprehension. That is, perhaps our subjects saw the questionnaire as a disguised measure of prejudice and sought to avoid appearing prejudiced by mitigating their prediction of group differences. This possibility seems unlikely for two reasons. First, the instructions emphasized an objective context for the estimates. Subjects knew the questions had correct answers that would be compared with their own as soon as the questionnaire was completed. Evaluation apprehension in this context should have been directed to making as accurate estimates as possible. Second, the amount or extremity of stereotyping was not systematically different for our six diverse groups of subjects. If subjects saw the questionnaire as a measure of prejudice, higher socioeconomic status and more sophisticated groups should have been more concerned to deny prejudice than lower status, less sophisticated groups. In fact, the college and MSW students and the caseworkers were stereotyping no less (significant stereotyping for 20 of 21 mean diagnostic ratios) than the union members and commercial-course high school students (significant stereotyping for 11 of 14 mean diagnostic ratios). We therefore conclude that dissembling of prejudice was not an important determinant of our results and that the tested stereotypes of black Americans are in fact relatively accurate and not exaggerated.

Theoretical Value of the Diagnostic Ratio: Stereotyping as Prediction and Concept

The proposed definition of stereotype places the study of stereotyping within the more formal and mathematical literature of the psychology of prediction. According to the new definition, the stereotype of any group of people—whether ethnic-national group or professional women or bus drivers or red-haired men or boys named Egbert—is composed of those attributes for which within-group predictions differ from base-rate predictions. From this definition, it follows that stereotypes are one kind of human prediction; results of stereotype studies should inform and complement studies of other kinds of prediction. Some beginning in this direction is made by Study 1, which indicates that people are not so generally unable to take account of base rate as Kahneman and Tversky (1973) have suggested.

An interesting implication of defining stereotypes by diagnostic ratios is that categories other than groups of people can be understood in the same terms. Consider what diagnostic ratios greater than 1.0 would mean for predictions about impersonal categories like house, car, or chair. For instance, p(four-legged/chair)/p(four-legged) is probably greater than 1.0, and the total of attributes having diagnostic ratios greater than 1.0 is very close to a positive definition of the concept of "chair." To know the attributes that distinguish any category, especially from similar categories, is to define the concept of that category. The study of stereotypes defined by diagnostic ratios is therefore a part of the more general study of human conceptual behavior.

Theoretical Value of the Diagnostic Ratio: Stereotyping and Attribution Theory

One aspect of human conceptual behavior that is much studied in social psychology is the attribution of causes of human behavior (Kelley, 1973). It seems that there must be a relation between attribution of causes and attribution of stereotype traits, and "A Bayesian Analysis of Attribution Theory" by Ajzen and Fishbein (1975) has paved the way for understanding this relation.

The Ajzen and Fishbein analysis may be expressed in the kind of conditional probabilities already used with stereotypes. If there are several possible hypotheses (H_1, H_2, H_3) to explain a behavior (B), the attributed cause of (B) is that H for which p(H/B)/p(H) is greatest. An example used by Ajzen and Fishbein is a wife who laughs at her husband's joke. She may have laughed because the joke was funny, because her husband is funny, or because she is flattering her husband. The attributed cause will follow whichever is greater: p(funny joke/laugh)/p(funny joke); p(funny husband/laugh)/p(funny husband); or p(flattering wife/laugh)/p(flattering wife).

It is important to note that causal attribution is said to depend on a ratio, p(H/B)/p(H), rather than only on the posterior probability, p(H/B). To emphasize this point, Ajzen and Fishbein consider the hypothesis that the wife laughed at the joke because she was a brunette. If she is indeed brunette, p(brunette/laugh) will be very high; but her hair color will not be seen as the cause of her laughter unless p(brunette/laugh)/p(brunette) is high. The argument is neatly parallel to the one offered in this article as to why stereotype traits are distinctive, not merely likely, for the stereotyped group.

From the vantage point provided by Ajzen and Fishbein's Bayesian analysis of attribution, it is possible to see the relation between stereotyping and causal attribution. The first is defined by a revision of behavioral (trait) probability on the evidence of group membership. The second is defined by a revision of causal probability on the evidence of behavior. But both revisions depend, as specified by Bayes' theorem, on the diagnostic ratio $p(H/D)/p(H)$ or equivalently, $p(D/H)/p(D)$. In stereotyping, H is the trait attribution and D is the group membership; in attribution theory, H is the causal attribution and D is the behavior.

Formally, even this distinction in viewpoint disappears. There is no formal distinction between (a) beginning with a category of person and inferring a category of behavior and (b) beginning with a category of behavior and inferring a (causal) category of person. Either way, one looks at an interdependence of attributes, and there is no reason other than convention for calling one attribute cause and the other effect, or one category the stereotyped group and the other the stereotype trait. One may as easily say that efficient people are stereotyped as German as that Germans are stereotyped as efficient. And one may as easily say that a person is a flatterer because she laughs at another's joke as say that she laughs because she is a flatterer. At the bottom of all these statements is an association of attributes such that, given one attribute, the probability of another is revised according to Bayes' theorem.

General Conclusion

The attractiveness of the diagnostic ratio as a stereotype measure is that it is the formal expression of the extent to which information about group membership affects trait predictions. Conceptually, this is not a new definition of stereotype. It is in fact the meaning of stereotype already implicit in a number of thoughtful discussions of stereotyping (Brown, 1965; LeVine & Campbell, 1972; Zawadzki, 1948). Made now explicit, this definition relates stereotyping to the psychology of prediction, to the study of conceptual behavior, and to attribution theory. Empirically, in relation to the familiar Katz and Braly (1933) checklist measure, the diagnostic ratio is something new. Study 2 in this article indicated that the diagnostic ratio measure, though related to the checklist, is also substantially different. It is always difficult to abandon a measure with as much research behind it as the checklist has, but the new measure promises to open many questions about stereotypes that cannot be investigated with the checklist. In Study 3, for instance, the new measure made possible what is, to our knowledge, the first systematic test of the "kernel of truth" hypothesis. The Katz and Braly measure has given us a literature of studies connected neither to one another nor to theory. Both conceptually and empirically, the diagnostic ratio measure offers a better alternative.

Reference Notes

1. McCauley, C., Stitt, C. L., & Greenberg, J. *Humans can be Bayesian: Intuitive predictions consistent with normative theory.* Unpublished manuscript, 1974.
2. McCauley, C., & Stitt, C. L. *Testing stereotype validity.* Paper presented at the meeting of the Eastern Psychological Association, New York City, April 1976.

References

Ajzen, I., & Fishbein, M. A Bayesian analysis of attribution processes. *Psychological Bulletin,* 1975, *82,* 261–277.

Brigham, J. C. Ethnic stereotypes, attitudes, and treatment of ethnic group members. (Doctoral dissertation, University of Colorado, 1969). *Dissertational Abstracts International,* 1969, *30,* (University Microfilms No. 70-5822)

Brigham, J. C. Ethnic stereotypes. *Psychological Bulletin,* 1971, *76,* 15–38. (a)

Brigham, J. C. Racial stereotypes, attitudes, and evaluations of and behavioral intentions toward Negroes and whites. *Sociometry,* 1971, *34,* 360–380. (b)

Brigham, J. C. Ethnic stereotypes and attitudes: A different mode of analysis. *Journal of Personality,* 1973, *41,* 206–233.

Brigham, J. C. Views of black and white children concerning the distribution of personality characteristics. *Journal of Personality*, 1974, *42*, 144–158.

Brown, R. *Social psychology*. New York: Free Press, 1965.

Gardner, R. C., Kirby, D. M., Gorospe, F. H., & Villamin, A. C. Ethnic stereotypes: An alternative assessment technique, the stereotype differential. *Journal of Social Psychology*, 1972, *87*, 259–267.

Jones, R. A., & Ashmore, R. D. The structure of intergroup perception: Categories and dimensions in views of ethnic groups and adjectives used in stereotype research. *Journal of Personality and Social Psychology*, 1973, *25*, 428–438.

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, *80*, 237–251.

Karlins, M., Coffman, T. L., & Walters, G. On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 1969, *13*, 1–16.

Katz, D., & Braly, K. W. Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, 1933, *28*, 280–290.

Kelley, H. H. The processes of causal attribution. *American Psychologist*, 1973, *28*, 107–128.

LeVine, R. A., & Campbell, D. T. *Ethnocentrism*. New York: Wiley, 1972.

U.S. Bureau of the Census. *Statistical abstract of the United States: 1975* (96th ed.). Washington, D.C.: U.S. Government Printing Office, 1975.

Zawadzki, B. Limitations of the scapegoat theory of prejudice. *Journal of Abnormal and Social Psychology*, 1948, *43*, 127–141.

Received June 30, 1977 ■