

---

# Ethnic and National Stereotypes: The Princeton Trilogy Revisited and Revised

**Stephanie Madon**

**Max Guyll**

*Iowa State University*

**Kathy Aboufadel**

**Eulices Montiel**

**Alison Smith**

**Polly Palumbo**

**Lee Jussim**

*Rutgers University*

---

*Three studies assessed changes in the content, consensus, and favorableness of 10 ethnic and national stereotypes by replicating and extending the Princeton trilogy. Results indicated that throughout the past 60 years, almost all of the ethnic and national stereotypes that were examined had changed in content, and more than half had changed in consensus. Most changes in consensus reflected increases rather than decreases, suggesting that modern members of stereotyped groups may confront stereotypes more frequently than did previous members of stereotyped groups. However, the damaging effects that consensual stereotypes can have on members of these groups may be tempered by the finding that most of the stereotypes became more favorable. These results are discussed in terms of changing social roles, intergroup contact, and stereotype accuracy.*

---

**W**hat do people believe about different social groups, and do these beliefs change from one generation to the next? Studies that address these kinds of questions once dominated social psychological research. From the 1930s to the 1950s, there was widespread interest in a variety of stereotype content issues, including the rigidity and inaccuracy of ethnic stereotypes (e.g., Clarke & Campbell, 1955; Katz & Braly, 1933; LaPierre, 1936). The Princeton trilogy, which includes Katz and Braly's (1933) original study and two replications (Gilbert, 1951; Karlins, Coffman, & Walters, 1969), exemplifies this early emphasis. These studies, which together spanned almost 40 years, examined the content (i.e., the attributes that people believe describe a social group) and consensus (i.e., the degree to which people agree on

which attributes describe a social group) of Princeton students' stereotypes.

In these studies, Princeton University students selected 5 personality traits, from a list of 84, that best described 10 ethnic and national groups. Katz and Braly (1933) found that Princeton students held distinct beliefs about each of the groups and that there was a high degree of consensus in their beliefs. Using Katz and Braly's method and attribute list, the replications assessed the extent to which the stereotypes changed in content and consensus over time. Gilbert (1951) found that the stereotypes had remained stable in content but had changed in consensus, whereas Karlins et al. (1969) found that the stereotypes had changed in content but had remained stable in consensus.

*Stereotype Content Research:*

*Decline and Revival*

The early emphasis on issues of stereotype content waned as the New Look in Perception gained popularity (Allport, 1954; Bruner, 1957). During this time, stereo-

---

**Authors' Note:** Earlier versions of this research were presented at the 68th annual conference of the Eastern Psychological Association in Washington, D.C. Additional tables not presented in this article are available at <http://psych-server.iastate.edu/faculty/madon> or on request. Portions of this research were performed while the first author was at Rutgers University. Correspondence regarding this article should be addressed to Stephanie Madon, Department of Psychology, W112 Lagomarcino Hall, Iowa State University, Ames, IA 50011-3180; e-mail: [madon@iastate.edu](mailto:madon@iastate.edu).

*PSPB*, Vol. 27 No. 8, August 2001 996-1010

© 2001 by the Society for Personality and Social Psychology, Inc.

type researchers directed their attention to the processes by which stereotypes shape social perception. This shift in emphasis led to some important findings. Stereotypes can bias impressions of individuals, produce self-fulfilling prophecies, and lead to discrimination and harassment (see Amicus Curiae Brief for the American Psychological Association, 1991; Fiske & Neuberg, 1990; Jussim, Eccles, & Madon, 1996, for reviews). Recently, however, there has been renewed interest in a variety of stereotype content issues. Consider the following examples. Contemporary perspectives on sex stereotypes suggest that hostile and benevolent sexist attitudes are related to how favorably people perceive women in general (Glick & Fiske, 1996, 1997). Research on stereotype accuracy has made a comeback after many years of decline, with researchers once again studying the correspondence between the content of stereotypes and criterion variables (e.g., Ashton & Esses, 1999; Judd & Park, 1993; Jussim et al., 1996; Madon et al., 1998). Research on prejudice indicates that most people are knowledgeable about the content of cultural stereotypes but that high-prejudice people personally hold those beliefs more than do low-prejudice people (Devine, 1989; Lepore & Brown, 1997).

The renewed interest in stereotype content may reflect the complimentary relationship between content and process. For example, the processes of discrimination and sexual harassment partially depend on the content of stereotypes. Men who hold hostile sexist attitudes may harass women because of the motive to dominate them, whereas men who hold benevolent sexist attitudes may harass women because of the motive to protect them (Fiske & Glick, 1995). Stereotype content also is linked to self-fulfilling prophecies because only inaccurate stereotypes can be self-fulfilling (Jussim et al., 1996). Finally, prejudice may stem from cultural stereotypes when controlled processes are inhibited but from personal stereotypes when controlled processes are not inhibited (Devine, 1989).

#### *Beyond the Princeton Trilogy*

There have been two comparatively recent follow-ups to the Princeton trilogy. In their examinations of the African American stereotype, both Devine and Elliot (1995) and Dovidio and Gaertner (1986) found that its content had changed substantially over time and that consensus about which of Katz and Braly's (1933) 84 attributes described African Americans had decreased. Devine and Elliot (1995) also supplemented Katz and Braly's (1933) attribute list with 9 new attributes and found low consensus for Katz and Braly's attributes but high consensus for the new attributes. This finding suggests that the attribute list used by the Princeton trilogy replications may now be outdated by virtue of omitting

current beliefs about the groups (see also Karlins et al., 1969).

In addition to making significant contributions to the assessment of stereotype change, recent studies of ethnic and national stereotypes also have raised several important questions. First, because the follow-ups to the Princeton trilogy (Devine & Elliot, 1995; Dovidio & Gaertner, 1986) focused exclusively on changes in the African American stereotype, it remains an open empirical question as to whether similar changes characterize other ethnic and national stereotypes. Second, few studies have examined changes in stereotype favorableness, and those that have (Devine & Elliot, 1995; Karlins et al., 1969; Seago, 1947; Sigall & Page, 1971), have examined the favorableness of attributes in general and not the favorableness of attributes in relation to the groups that they supposedly describe. This can be problematic because attributes can have different meanings and, therefore, different favorableness, depending on group membership (Kunda, Sinclair, & Griffin, 1997). Finally, because nearly every study examining stereotype change has assessed European American's stereotypes about other ethnic groups, little is known about the stereotypes endorsed by other groups.

#### *Overview of Current Research*

The current research includes three studies that examined the extent to which ethnic and national stereotypes have changed in content, consensus, and favorableness over time. This research addresses several important issues raised by previous work on these topics. Specifically, in addition to fully replicating the Princeton trilogy, this research also updates Katz and Braly's (1933) attribute list with more than 300 new attributes. This enables an assessment of whether Katz and Braly's original attribute list has become outdated and whether its use might, therefore, mask real changes in the stereotypes. This work also examines changes in stereotype favorableness by assessing the favorableness of stereotypic attributes in relation to the groups that those attributes are perceived to describe. Finally, this research investigates these issues for both European American and non-European American samples.

#### STUDY 1: REPLICATION OF THE PRINCETON TRILOGY

The primary purpose of this research was to examine whether ethnic and national stereotypes have changed over time. We began our investigation of this issue by fully replicating the Princeton trilogy. This replication provided a baseline assessment of the current stereotypes that perceivers endorse when they are provided with Katz and Braly's (1933) original attribute list.

## Method

### PARTICIPANTS

Participants were 102 undergraduates enrolled in introductory psychology courses. There were 61 White participants in the European American sample (34 women and 27 men) and 41 participants in the non-European American sample (27 women and 14 men; 2 African Americans, 22 Asians, 8 Latino/as, 1 Native American, and 8 participants who coded their ethnicity as "Other").

### QUESTIONNAIRE AND PROCEDURES

The procedures and materials replicated those used by the Princeton trilogy studies. Participants completed a questionnaire that listed Katz and Braly's (1933) 84 attributes. The questionnaire first instructed participants to write down as many attributes as necessary to characterize each of these 10 groups: Germans, Italians, African Americans, Irish, English, Jews, Americans, Chinese, Japanese, and Turks. The questionnaire then instructed participants to mark an "X" by the five attributes that were most typical of each group.

## Results and Discussion

### STEREOTYPE CONTENT

Analyses addressed three questions related to the content of ethnic and national stereotypes: (a) What is the content of current stereotypes? (b) Do European and non-European Americans endorse similar stereotypes? and (c) Has the content of the stereotypes changed over time?

*Content of current stereotypes.* Analyses assessed the content of the stereotypes with the method used by the Princeton trilogy. These analyses were performed separately for the European and non-European American samples. For each of the 10 groups, frequency analyses identified the percentage of participants who endorsed each attribute as one of the five most characteristic. We defined the content of a stereotype as the 10 most frequently endorsed attributes. Results showed that the European American sample most often perceived Germans as intelligent, Italians as loyal to family ties, African Americans as musical, Irish as happy-go-lucky, English as sophisticated, Jews as very religious, Americans as materialistic, Chinese and Japanese as intelligent, and Turks as extremely nationalistic. The non-European American sample most often perceived Germans as industrious, Italians as loyal to family ties, African Americans as loud, Irish as happy-go-lucky, English as conservative, Jews as very religious, Americans as materialistic, Chinese as loyal to family ties, Japanese as intelligent, and Turks as extremely nationalistic. Tables 1 and 2 present

the American and African American stereotypes. Tables presenting the other stereotypes are available at <http://psych-server.iastate.edu/faculty/madon>.

*Differences between stereotypes endorsed by European and non-European Americans.* The Kappa statistic (Cohen, 1960) assessed the extent to which the European and non-European American samples endorsed similar stereotypes by testing whether these samples agreed beyond chance with regard to the attributes that they selected as descriptive of the 10 groups. Kappa equals zero when observed agreement is no better than chance and equals one when observed agreement is perfect. We used guidelines from Landis and Koch (1977) to interpret the amount of agreement reflected by Kappa's magnitude.

Results indicated that agreement between the attributes endorsed as descriptive of the groups by our European and non-European American samples always exceeded chance, all  $ps < .001$ . Agreement was moderate for the Turkish stereotype ( $K = .55$ ); substantial for the German, Italian, African American, Irish, American, Chinese, and Japanese stereotypes ( $.66 \geq$  all  $Ks \leq .79$ ); and almost perfect for the English and Jewish stereotypes (both  $Ks = .89$ ). These results indicate remarkable similarity between the stereotypes endorsed by the European and non-European American samples. However, this similarity may reflect the use of an outdated attribute list. The samples may have endorsed old and, therefore, similar stereotypes for a lack of better alternatives, making their stereotypes appear more similar than they actually were.

*Changes in stereotype content over time.* The Kappa statistic also examined changes in the content of the stereotypes over time by testing whether our European American sample and the Princeton trilogy samples agreed beyond chance with regard to the attributes that they endorsed as descriptive of the groups.<sup>1</sup> Results indicated that change was greatest for the African American stereotype. Agreement between the attributes endorsed as descriptive of African Americans by our European American sample and the three Princeton trilogy samples never exceeded chance (all  $Ks \leq .13$ ; all  $ps > .05$ ). This result replicates the findings of the Princeton trilogy follow-ups (Devine & Elliot, 1995; Dovidio & Gaertner, 1986), which found that the content of the African American stereotype changed over time. However, because the follow-ups had only examined the African American stereotype, neither could determine whether other stereotypes had undergone similar changes over time. One goal of our research was to address this issue.

Results suggested that the other ethnic and national stereotypes had changed less than did the African American stereotype. Agreement between our sample and the

TABLE 1: American Stereotypes

Attribute	Katz and Braly (1933)	Gilbert (1951)	Karlins, Coffman, and Walters (1969)	Study 1		Study 2	
				EA	NEA	EA	NEA
Industrious	48.5	30.0	23.0	23.2	15.8	12.9	11.8
Intelligent	47.5	32.0	20.0	17.9	26.3	6.5	11.8
Materialistic	33.3	37.0	67.0	53.6	36.8	25.8	29.4
Ambitious	33.3	21.0	42.0	10.7	10.5	16.1	5.9
Progressive	27.3	5.0	17.0	10.7	7.9	6.5	5.9
Pleasure loving	26.3	27.0	28.0	26.8	26.3	16.1	12.5
Alert	23.2	7.0	7.0	1.8	2.6	12.9	5.9
Efficient	21.2	9.0	15.0	3.6	2.6	6.5	5.9
Aggressive	20.2	8.0	15.0	12.5	15.8	19.4	11.8
Straightforward	19.2	—	9.0	8.9	7.9	12.9	0.0
Practical	19.2	—	12.0	1.8	5.3	9.7	0.0
Sportsmanlike	19.2	—	9.0	7.1	15.8	9.7	11.8
Individualistic	—	26.0	15.0	28.6	31.6	22.6	5.9
Conventional	—	—	17.0	5.4	5.3	9.7	0.0
Scientifically minded	—	—	15.0	7.1	7.9	9.7	12.5
Ostentatious	—	—	15.0	8.9	7.9	6.9	0.0
Lazy	—	—	—	30.4	15.8	22.6	23.5
Extremely nationalistic	—	—	—	19.6	15.8	9.7	17.7
Ignorant	—	—	—	16.1	13.2	9.7	0.0
Impulsive	—	—	—	16.1	7.9	22.6	11.8
Arrogant	—	—	—	14.3	15.8	19.4	5.9
Rude	—	—	—	14.3	7.9	16.1	11.8
Diverse <sup>a</sup>	—	—	—	—	—	64.5	41.2
Democratic <sup>a</sup>	—	—	—	—	—	32.3	23.5
Listen to a lot of music <sup>a</sup>	—	—	—	—	—	32.3	17.7
Flirtatious <sup>a</sup>	—	—	—	—	—	30.0	6.3
Competitive <sup>a</sup>	—	—	—	—	—	29.0	17.7
Loud	—	—	—	—	—	29.0	5.9
Outspoken <sup>a</sup>	—	—	—	—	—	29.0	5.9
Stubborn	—	—	—	—	—	29.0	5.9
Interests wide <sup>a</sup>	—	—	—	—	—	26.7	6.3
Adventurous <sup>a</sup>	—	—	—	—	—	25.8	12.5
Boastful	—	—	—	—	—	25.8	5.9
Cool <sup>a</sup>	—	—	—	—	—	25.8	0.0
Hard-headed <sup>a</sup>	—	—	—	—	—	25.8	5.9
Independent <sup>a</sup>	—	—	—	—	—	25.8	5.9
Leaders <sup>a</sup>	—	—	—	—	—	25.8	11.8
Liberal <sup>a</sup>	—	—	—	—	—	25.8	11.8
Opinionated <sup>a</sup>	—	—	—	—	—	25.8	18.8
Rebellious <sup>a</sup>	—	—	—	—	—	25.8	23.5
Prejudiced <sup>a</sup>	—	—	—	—	—	19.4	35.3
Superficial <sup>a</sup>	—	—	—	—	—	12.9	29.4
Emotional <sup>a</sup>	—	—	—	—	—	9.7	25.0
Complaining <sup>a</sup>	—	—	—	—	—	16.1	23.5
Cruel	—	—	—	—	—	12.9	23.5
Greedy <sup>a</sup>	—	—	—	—	—	19.4	23.5
Patriotic <sup>a</sup>	—	—	—	—	—	16.1	23.5
Politically active <sup>a</sup>	—	—	—	—	—	6.5	23.5
Proud <sup>a</sup>	—	—	—	—	—	19.4	23.5
Racists <sup>a</sup>	—	—	—	—	—	16.1	23.5
Show-offs <sup>a</sup>	—	—	—	—	—	22.6	23.5
Spoiled <sup>a</sup>	—	—	—	—	—	22.6	23.5

NOTE: EA refers to the European American samples; NEA refers to the non-European American samples. For the Princeton trilogy and Study 1, values reflect the percentage of participants who endorsed an attribute as one of the five most characteristic of Americans. For Study 2, values reflect the percentage of participants who endorsed an attribute as "much more characteristic of Americans than other people." Attributes without superscripts are from Katz and Braly's attribute list. Values in italics correspond to the 10 most frequently endorsed attributes for Studies 1 and 2. In the case of ties, attributes with the same percentages are listed.

a. Identifies attributes added in Study 2.

TABLE 2: African American Stereotypes

Attribute	Katz and Braly (1933)	Gilbert (1951)	Karlins, Coffman, and Walters (1969)	Study 1		Study 2	
				EA	NEA	EA	NEA
Superstitious	84.0	41.0	13.0	1.7	5.3	0.0	4.2
Lazy	75.0	31.0	26.0	12.1	10.5	0.0	4.2
Happy-go-lucky	38.0	17.0	27.0	0.0	7.9	0.0	4.2
Ignorant	38.0	24.0	11.0	10.3	13.2	3.5	0.0
Musical	26.0	33.0	47.0	27.6	39.5	3.5	33.3
Ostentatious	26.0	11.0	25.0	8.6	5.3	0.0	4.6
Very religious	24.0	17.0	8.0	19.0	13.2	6.9	16.7
Stupid	22.0	10.0	4.0	1.7	0.0	0.0	0.0
Physically dirty	17.0	—	3.0	0.0	0.0	3.5	0.0
Naive	14.0	—	4.0	1.7	0.0	0.0	0.0
Slovenly	13.0	—	5.0	3.5	7.9	0.0	0.0
Unreliable	12.0	—	6.0	5.2	5.3	0.0	0.0
Pleasure loving	—	19.0	26.0	5.2	0.0	0.0	25.0
Sensitive	—	—	17.0	5.2	0.0	0.0	4.2
Gregarious	—	—	17.0	6.9	5.3	0.0	8.7
Talkative	—	—	14.0	10.3	13.2	6.9	20.8
Imitative	—	—	13.0	1.7	0.0	3.5	8.3
Loyal to family ties	—	—	—	22.4	13.2	3.5	12.5
Loud	—	—	—	20.7	52.6	6.9	20.8
Tradition loving	—	—	—	20.7	7.9	3.5	8.3
Aggressive	—	—	—	15.5	18.4	0.0	16.7
Artistic	—	—	—	13.8	5.3	3.5	8.3
Quick tempered	—	—	—	13.8	15.8	6.9	16.7
Revengeful	—	—	—	13.8	5.3	0.0	8.3
Rude	—	—	—	13.8	26.3	0.0	20.8
Quarrelsome	—	—	—	8.6	18.4	0.0	20.8
Sportsmanlike	—	—	—	12.1	15.8	6.9	25.0
Intelligent	—	—	—	3.5	13.2	0.0	4.2
Materialistic	—	—	—	8.6	13.2	0.0	8.3
Passionate	—	—	—	8.6	13.2	0.0	16.7
Straightforward	—	—	—	12.1	13.2	6.9	21.7
Listen to a lot of music <sup>a</sup>	—	—	—	—	—	17.2	25.0
Noisy <sup>a</sup>	—	—	—	—	—	17.2	25.0
Athletic <sup>a</sup>	—	—	—	—	—	13.8	26.1
Have an attitude <sup>a</sup>	—	—	—	—	—	13.8	25.0
Prejudiced <sup>a</sup>	—	—	—	—	—	13.8	20.8
Sing and dance well <sup>a</sup>	—	—	—	—	—	13.8	20.8
Cultural <sup>a</sup>	—	—	—	—	—	10.7	8.3
Democratic <sup>a</sup>	—	—	—	—	—	10.3	4.2
Angry <sup>a</sup>	—	—	—	—	—	10.3	20.8
Masculine <sup>a</sup>	—	—	—	—	—	10.3	8.3
Opinionated <sup>a</sup>	—	—	—	—	—	10.3	29.2
Outspoken <sup>a</sup>	—	—	—	—	—	10.3	12.5
Tough <sup>a</sup>	—	—	—	—	—	6.9	33.3
Humorous <sup>a</sup>	—	—	—	—	—	3.5	30.4
Rebellious <sup>a</sup>	—	—	—	—	—	0.0	29.2
Active <sup>a</sup>	—	—	—	—	—	0.0	25.0
Bossy <sup>a</sup>	—	—	—	—	—	0.0	25.0
Energetic <sup>a</sup>	—	—	—	—	—	0.0	25.0
Outgoing <sup>a</sup>	—	—	—	—	—	0.0	25.0
Proud <sup>a</sup>	—	—	—	—	—	3.5	25.0
Show-offs <sup>a</sup>	—	—	—	—	—	0.0	25.0
Strong <sup>a</sup>	—	—	—	—	—	3.5	25.0

NOTE: EA refers to the European American samples; NEA refers to the non-European American samples. For the Princeton trilogy and Study 1, values reflect the percentage of participants who endorsed an attribute as one of the five most characteristic of African Americans. For Study 2, values reflect the percentage of participants who endorsed an attribute as "much more characteristic of African Americans than other people." Attributes without superscripts are from Katz and Braly's attribute list. Values in italics correspond to the 10 most frequently endorsed attributes for Studies 1 and 2. In the case of ties, attributes with the same percentages are listed.

three Princeton trilogy samples always exceeded chance for the German, Italian, Irish, English, Jewish, American, Chinese, and Turkish stereotypes (all  $ps \leq .03$ ). Agreement was fair for the American stereotype ( $.27 \leq Ks \leq .37$ ) and moderate to substantial for the German ( $.55 \leq Ks \leq .75$ ), Italian ( $.43 \leq Ks \leq .66$ ), Irish ( $.55 \leq Ks \leq .69$ ), English ( $.55 \leq Ks \leq .66$ ), Jewish ( $.59 \leq Ks \leq .66$ ), Chinese ( $.51 \leq Ks \leq .72$ ), and Turkish ( $.43 \leq Ks \leq .69$ ) stereotypes. Agreement between our sample and Gilbert's (1951) and Karlins et al.'s (1969) samples also exceeded chance for the Japanese stereotype ( $ps < .001$ ), showing moderate agreement in both cases ( $.45 \geq Ks \leq .55$ ). In contrast, agreement between our sample and Katz and Braly's (1933) sample did not exceed chance for the Japanese stereotype ( $K = .13, p > .05$ ).

In sum, these results suggest that the content of most of the stereotypes did not change substantially over time. However, because we assessed stereotype change with a possibly outdated attribute list, these results should be interpreted with caution. Participants may have endorsed old stereotypes because that was all that was available on the attribute list. If this occurred, then the degree of stereotype change that we report may underestimate the degree of change that the stereotypes have actually undergone.

#### STEREOTYPE CONSENSUS

Two procedures were employed to examine stereotype consensus. The first was based on distinctiveness scores, as developed by Katz and Braly (1933). Distinctiveness scores were calculated separately for the European and non-European American samples. For each stereotype, attributes were ordered from high to low according to the frequency with which each was endorsed as one of the five most characteristic. Beginning with the most frequently endorsed attribute, these frequencies were summed until the cumulative frequency equaled one half the number of total endorsements provided by the European American sample ( $152.5 = [61 \text{ participants} \times 5 \text{ endorsements per participant}] / 2$ ) and the non-European American sample ( $102.5 = [41 \times 5] / 2$ ).

Lower distinctiveness scores indicate greater consensus. For the European American sample, distinctiveness scores were as follows: Germans, 10.4; Italians, 6.7; African Americans, 17.5; Irish, 10.4; English, 8.6; Jews, 8.6; Americans, 11.9; Chinese, 7.1; Japanese, 7.5; and Turks, 40.3. For the non-European American sample, distinctiveness scores were as follows: Germans, 13.9; Italians, 10.4; African Americans, 13.3; Irish, 15.2; English, 13.5; Jews, 7.9; Americans, 15.5; Chinese, 9.3; Japanese, 9.9; and Turks, 23.8.

Because the distinctiveness score measure of consensus yields only a single value for each stereotype, it does

not have an associated estimate of variability. This precludes the ability to analyze these data with inferential statistics to test for significant differences. Therefore, we also measured consensus with the percentages corresponding to the most frequently endorsed attributes in each stereotype (e.g., see Tables 1 and 2, Study 1).

These percentages vary across attributes within a stereotype. For example, Table 1 shows that there was greater consensus among our European American sample that Americans are materialistic (53.6%) than that they are rude (14.3%). Because of this variability, this percentage-based measure of consensus can be analyzed with inferential statistics. Nonparametric<sup>2</sup> statistics based on these percentages addressed three questions related to consensus: (a) Do the current stereotypes vary in consensus? (b) Do the stereotypes endorsed by European and non-European Americans differ in consensus? and (c) Have there been changes in stereotype consensus over time?

*Consensus differences between current stereotypes.* Kruskal-Wallis analyses tested whether the current stereotypes varied in consensus. The Kruskal-Wallis is the nonparametric equivalent of the one-way analysis of variance; it tests for a difference between three or more medians (see Gibbons & Chakraborti, 1992). Separate Kruskal-Wallis analyses were performed for the European and non-European American samples. Although analyses revealed that significant differences in consensus existed between the 10 stereotypes for the European American sample,  $\chi^2(9) = 17.71, p = .04$ , subsequent contrasts that controlled for family-wise Type I error ( $\alpha_{\text{familywise}} = .05$ ) failed to yield any significant pairwise differences. Analyses did not reveal significant differences in consensus between the 10 stereotypes for the non-European American sample,  $\chi^2(9) = 7.85, p = .55$ .

*Consensus differences between European and non-European Americans.* The Mann-Whitney  $U$  tested for consensus differences between the stereotypes endorsed by the European and non-European American samples. The Mann-Whitney  $U$  is the nonparametric equivalent of the  $t$  test; it tests for differences between two medians (see Gibbons & Chakraborti, 1992). Ten Mann-Whitney  $U$  tests were performed, one for each stereotype. Results showed no significant differences in consensus between the samples for the German, Italian, African American, Irish, English, Jewish, American, Chinese, and Japanese stereotypes ( $Us \geq 32, ps \geq .17$ ). The non-European American sample did, however, endorse a significantly more consensual stereotype about Turks than did the European American sample ( $U = 24, p = .05$ ). These results suggest that few differences existed between the European and non-European American samples with regard to stereotype consensus. However, as was the case with

content, this apparent similarity may partly reflect the use of an outdated attribute list. The samples may have endorsed old stereotypes with similar frequency because the list excluded current beliefs about the 10 groups.

*Changes in consensus over time.* Analyses next examined whether stereotype consensus has changed over time. Three Mann-Whitney  $U$  tests were performed for each stereotype. These analyses compared the consensus of the stereotype endorsed by our European American sample to the consensus of the stereotype endorsed by the three Princeton trilogy samples (see Note 1). Bonferroni corrections controlled for Type I error arising from these three comparisons by reducing the significance level to  $p \leq .017$ . Results indicated that our sample held a less consensual stereotype about African Americans than did Katz and Braly's (1933) sample ( $U = 16, p = .01$ ) and more consensual stereotypes about Chinese, Japanese, and Turks than did Gilbert's (1951) sample (all  $U_s \leq 7, p_s \leq .01$ ). Comparisons between our sample and those of the Princeton trilogy with respect to consensus were not significant for any other stereotypes (all  $U_s \geq 12, p_s \geq .11$ ). These results suggest that most of the stereotypes remained stable in consensus over time. However, these results may have been affected by the use of an outdated attribute list. Participants may have been knowledgeable about old stereotypes and may have exhibited levels of consensus similar to those of previous samples when provided with an attribute list that omitted current stereotypes. This type of responding may have made the consensus of the current stereotypes appear similar to the consensus of the stereotypes reported by the Princeton trilogy.

#### STUDY 2: 10 STEREOTYPES REVISITED WITH AN UPDATED ATTRIBUTE LIST

Study 1 suggested that over time most of the stereotypes under investigation remained stable in content and consensus. However, this stability may have been more apparent than real, arising out of the use of an outdated attribute list. Therefore, in Study 2, we updated Katz and Braly's (1933) attribute list and then reassessed the content and consensus of the stereotypes and the extent to which those stereotypes have changed over time. In addition to including Katz and Braly's original attribute list, the updated list also included personality traits from Gough and Heilbrun's (1983) adjective checklist<sup>3</sup> and free responses obtained from a preliminary study.

##### *Preliminary Study*

The study included 41 participants who listed up to three attributes that they believed described each of the 10 groups. This procedure yielded between 22 and

94 free responses for each group. The free responses were reduced by the following procedures. First, attributes listed only once were discarded to exclude idiosyncratic beliefs. Second, attributes that overlapped with Katz and Braly's (1933) attribute list or with Gough and Heilbrun's (1983) adjective checklist were discarded to eliminate redundancy. Third, two judges met and by consensus discarded attributes that did not reflect personal attributes (e.g., "green" for Irish, "Hanukkah" for Jews). Fourth, the judges identified attributes that were synonymous in meaning and selected one attribute or attribute-phrase to represent those attributes. These procedures yielded a total of 68 free responses.

#### MAIN STUDY 2

##### *Method*

##### *PARTICIPANTS*

Participants were 492 undergraduates enrolled in introductory psychology courses. There were 304 White participants in the European American sample (194 women and 110 men) and 188 participants in the non-European American sample (119 women and 69 men; 30 African Americans, 101 Asians, 38 Latino/as, and 19 participants who coded their ethnicity as "Other").

##### *PROCEDURES AND QUESTIONNAIRE*

Participants completed a questionnaire that included one group label and 406 attributes. The attributes consisted of the 84 attributes used in the Princeton trilogy, 68 free responses from the preliminary study, and 254 personality traits from Gough and Heilbrun's (1983) adjective checklist.<sup>4</sup> Participants rated the extent to which each attribute described the group on a 5-point scale (where 1 = *much less characteristic of "GROUP" than other people*, 2 = *somewhat less characteristic of "GROUP" than other people*, 3 = *equally as characteristic of "GROUP" as of other people*, 4 = *somewhat more characteristic of "GROUP" than other people*, and 5 = *much more characteristic of "GROUP" than other people*).

##### *Results and Discussion*

We performed two sets of analyses that paralleled those performed in Study 1. The first set of analyses examined the content of ethnic and national stereotypes; the second set of analyses examined the consensus of the stereotypes.

##### *STEREOTYPE CONTENT*

The first set of analyses reexamined the following questions related to content: (a) What is the content of current stereotypes? (b) Do European and non-European Americans endorse similar stereotypes? and

(c) Has the content of the stereotypes changed over time?

*Content of current stereotypes.* Following procedures similar to those of the Princeton trilogy, frequency analyses assessed the content of the stereotypes by identifying the attributes that participants most often endorsed as “much more characteristic” of each group. This was done separately for the European and non-European American samples. We defined the content of each stereotype as including the 10 most frequently endorsed attributes. The European American sample most often perceived Germans as liking to drink beer, Italians as loyal to family ties, African Americans as listening to a lot of music and noisy, Irish as liking to drink beer, English as competitive, Jews as very religious and wealthy, Americans as diverse, Chinese as disciplined, Japanese as scientifically minded and disciplined, and Turks as very religious. The non-European American sample most often perceived Germans as liking to drink beer; Italians as macho; African Americans as musical and tough; Irish as liking to drink beer; English as egotistical, proud, and liberal; Jews as proud; Americans as diverse; Chinese and Japanese as loyal to family ties; and Turks as cultural. Tables 1 and 2 present the American and African American stereotypes. Tables presenting the other stereotypes are available at <http://psych-server.iastate.edu/faculty/madon>.

*Differences in the stereotypes endorsed by European and non-European Americans.* The Kappa statistic assessed whether the European and non-European American samples endorsed similar stereotypes. Results indicated that the samples agreed beyond chance for all 10 groups (all  $ps < .01$ ). Agreement was slight for the Italian ( $K = .18$ ) and American ( $K = .20$ ) stereotypes; fair for the African American ( $K = .40$ ) and English ( $K = .35$ ) stereotypes; moderate for the German, Irish, Jewish, and Turkish stereotypes ( $.42 \geq$  all  $Ks \leq .59$ ); and substantial for the Chinese ( $K = .74$ ) and Japanese ( $K = .69$ ) stereotypes. These results show that the European and non-European American samples endorsed similar stereotypes overall. However, the degree to which they did so was less here than it was in Study 1, where agreement ranged from moderate to almost perfect. Differences in agreement across the studies may reflect differences in the attribute lists. Participants in Study 1 may have been constrained to endorse old and, therefore, similar stereotypes because they lacked attributes that are more relevant to contemporary stereotypes. In contrast, participants in Study 2 may not have faced this constraint because they received an updated attribute list.

Subsequent analyses explored these differences by comparing the European and non-European American samples' stereotypes on an attribute-by-attribute basis.

Participants' ratings of how much each attribute described the groups on the 5-point scale constituted the dependent variable. Each analysis included both the attributes that the European American sample endorsed as descriptive of the group and the attributes that the non-European American sample endorsed as descriptive of the group. Only those attributes identified as part of the current stereotypes in Study 2 were included in these analyses (e.g., see Tables 1 and 2). Data were analyzed with 10 separate repeated-measures ANOVAs, 1 for each stereotype. The between-subjects factor was ethnicity of participant (European vs. non-European American). The attributes comprising a stereotype constituted the levels of the within-subject factor. Because of different amounts of overlap between the European and non-European American samples' stereotypes, and because of different numbers of attributes comprising the stereotypes (due to ties), the within-subject factor varied from 18 attributes (Jews, Chinese, and Japanese) to 38 attributes (Germans). A significant interaction between ethnicity of participant and the within-subject factor would indicate that there was a significant difference in the extent to which the European and non-European American samples believed that individual attributes in the stereotype described the group. When results yielded a significant interaction, independent sample  $t$  tests identified the precise attributes for which these differences existed. Because these analyses were exploratory in nature, we did not adjust significance values for multiple comparisons.

Results yielded a significant interaction for the Italian, African American, English, Jewish, American, and Chinese stereotypes (all  $Fs \geq 6.25$ ,  $ps \leq .02$ ). In comparison to the European American sample,  $t$  tests indicated that the non-European American sample rated (a) Italians higher on demanding; (b) African Americans higher on sing and dance well, masculine, musical, tough, strong, pleasure loving, energetic, active, outgoing, and proud; (c) English higher on liberal, leaders, and dominant and lower on fussy; (d) Jews higher on patriotic; (e) Americans higher on prejudiced and lower on loud, stubborn, and outspoken; and (f) Chinese lower on disciplined (all  $ts \geq 2.07$ ,  $ps \leq .04$ ). Thus, even though the European and non-European American samples exhibited significant agreement in the content of their stereotypes overall, important differences nonetheless emerged with regard to their beliefs about particular attributes.

*Changes in stereotype content over time.* The Kappa statistic examined changes in the content of the stereotypes over time by testing whether our European American sample and the Princeton trilogy samples agreed beyond chance with regard to the attributes that they endorsed as descriptive of the groups (see Note 1). Agreement failed to exceed chance for the German, Italian, African



American, English, Jewish, American, Chinese, Japanese, and Turkish stereotypes, suggesting that the contents of these stereotypes have changed over time (all  $Ks \leq .21$ ,  $ps > .05$ ). Agreement between our sample and the three Princeton trilogy samples did exceed chance for the Irish stereotype, suggesting that its content was more stable by comparison ( $.22 \leq Ks \leq .37$ ,  $ps < .05$ ). The magnitude of the Kappas for the Irish stereotype reflected fair agreement in comparison to each study in the Princeton trilogy. The lack of agreement across time found in this study stands in contrast to the results of Study 1, the results for which suggested that only the content of the African American stereotype had changed over time.

#### STEREOTYPE CONSENSUS

The second set of analyses reexamined the following questions related to consensus: (a) Do current stereotypes vary in consensus? (b) Do the stereotypes endorsed by European and non-European Americans differ in consensus? And (c) Have there been changes in stereotype consensus over time? Analyses addressed these questions with the percentages obtained in Study 2 and the Princeton trilogy (e.g., see Tables 1 and 2).

*Consensus differences between current stereotypes.* Kruskal-Wallis analyses examined differences in the consensus between the current stereotypes. Analyses were performed separately for the European and non-European American samples. Results indicated significant differences in consensus among the 10 stereotypes for both of our samples, both  $\chi^2s(9) \geq 61.62$ ,  $ps < .001$ . Pairwise comparisons that controlled for family-wise Type I error ( $\alpha_{\text{family wise}} = .05$ ) further revealed that among the European American sample, the Chinese stereotype was significantly more consensual than the Jewish, German, English, and African American stereotypes; the American and Japanese stereotypes were significantly more consensual than the German, English, and African American stereotype; and the Italian and Irish stereotypes were significantly more consensual than the African American stereotype. Among the non-European American sample, the Japanese stereotype was significantly more consensual than the English, Irish, African American, Jewish, and American stereotypes; the Chinese stereotype was significantly more consensual than the Irish, African American, Jewish, and American stereotypes; and the Italian stereotype was significantly more consensual than the African American, Jewish, and American stereotypes.

*Consensus differences between European and non-European Americans.* Ten Mann-Whitney  $U$  tests, one for each stereotype, tested for consensus differences in the stereotypes endorsed by the European and non-

European American samples. Results indicated that the non-European American sample endorsed significantly more consensual stereotypes about Germans, Italians, African Americans, English, Jews, Chinese, Japanese, and Turks than did the European American sample (all  $Us \leq 18$ ,  $ps \leq .01$ ). The European American sample endorsed a significantly more consensual stereotype about Americans than did the non-European American sample ( $U = 30$ ,  $p = .04$ ). There were no differences in consensus for the Irish stereotype ( $U = 41$ ,  $p = .49$ ). These results indicate larger differences in consensus between the European and non-European American samples than did Study 1, thus providing more evidence that Study 2 was better able to detect real differences between samples' stereotypes because it used an updated attribute list.

*Changes in consensus over time.* Analyses also examined whether the stereotypes have changed in consensus over time. For each stereotype, three Mann-Whitney  $U$  tests compared the consensus of each stereotype endorsed by our European American sample to the consensus of the stereotype endorsed by each of the Princeton trilogy samples (see Note 1). Bonferroni corrections controlled for Type I error arising from these three comparisons by reducing the significance level to  $p \leq .017$ . Results both replicated and extended the findings of Study 1. Replicating the findings of Study 1, our sample held a less consensual stereotype about African Americans than did Katz and Braly's (1933) sample ( $U = 16$ ,  $p = .001$ ) and a more consensual stereotype about Chinese, Japanese, and Turks than did Gilbert's (1951) sample (all  $Us \leq 2$ ,  $ps \leq .002$ ). Findings of Study 2 also found that (a) in comparison to Katz and Braly's and Karlins et al.'s (1969) samples, our sample held a significantly more consensual stereotype about the Chinese (both  $Us = 12$ ,  $ps = .001$ ) and a significantly less consensual stereotype about the English (both  $Us \leq 9$ ,  $ps < .01$ ), and (b) in comparison to Karlins et al.'s sample, our sample held a significantly more consensual stereotype about Americans and Turks and a significantly less consensual stereotype about African Americans (all  $Us \leq 27$ ,  $ps \leq .01$ ). There were no other significant differences in consensus (all  $Us \geq 10$ ,  $ps \geq .02$ ). These results suggest that the findings of Study 1 may have underestimated how much ethnic and national stereotypes have changed in consensus over time and that the materials and procedures of Study 2 enabled a more sensitive analysis of these changes.

#### ALTERNATIVE EXPLANATIONS FOR THE PATTERNS OF CHANGE

The findings of Study 2 indicated that current ethnic and national stereotypes differed from those reported by the Princeton trilogy. This result led us to conclude that

the stereotypes have changed over time; however, these differences could have stemmed from procedural differences between the studies. Participants in the Princeton trilogy were presented with 84 attributes, made categorical responses as to whether a particular attribute described a group, and evaluated all 10 groups. In contrast, participants in Study 2 of this article were presented with 406 attributes, used a rating scale to indicate the extent to which each attribute described a group, and evaluated only one group. These procedural differences raise several alternative explanations for our findings that could potentially undermine our conclusion that the stereotypes have changed over time.

First, people may not respond to a longer list of attributes the same way that they respond to a shorter list of attributes. This may be especially true with respect to consensus. People may show less consensus in their responses when they have more attributes from which to choose. However, in comparison to the Princeton trilogy, most of the consensus changes found in Study 2 reflected increases, not decreases, in consensus. This means that if the use of a longer attribute list in Study 2 did tend to decrease consensus among participants' responses, then the bulk of the stereotypes that we examined may have become even more consensual over time than our results suggest.

Second, the longer list of attributes used in Study 2 also has implications for results regarding changes in stereotype content. Participants in Study 2 endorsed different attributes than did participants in the Princeton trilogy, indicating that using an updated attribute list did influence participants' responses. However, a fundamental goal of our research was to examine whether the content of the stereotypes changed over time and whether Katz and Braly's (1933) attribute list masked these changes by virtue of being outdated. Updating Katz and Braly's attribute list, therefore, was a necessary evil. Admittedly, we could have made the lists more comparable in number by adding fewer attributes to the updated attribute list. However, we are not aware of any research documenting how many new attributes would have been too many. Because of the ambiguity involved in deciding on the "correct" number of new attributes to add to the list, we chose to err on the side of inclusiveness so that the attribute list was more likely to contain the most relevant stereotypic attributes of each group.

Third, procedural changes in the judgment task raise the concern that people may respond differently when making categorical versus rating scale judgments. However, previous research indicates that it makes little difference whether researchers assess stereotypes with checklists, percentage estimates, diagnostic ratios, stereotype differentials, or rating scales, because these vari-

ous stereotype assessment procedures tend to produce highly intercorrelated results (Stangor & Lange, 1994).

A final concern involves the notion of comparison. Participants in the Princeton trilogy (who evaluated all 10 groups) may have assumed that their responses were to be made in comparison to other groups, whereas participants in Study 2 (who evaluated only one group) may not have made this assumption. We believe that this procedural difference was unlikely to have had a major effect on the findings because, as detailed in the Method section, the response options of Study 2 incorporated the notion of comparison by having participants rate each attribute in comparison to other people. Furthermore, people intuitively use a standard of comparison when evaluating social groups, even when that standard of comparison is not explicitly indicated (McCauley & Stitt, 1978).

In summary, the foregoing discussion argues against the possible influence of procedural differences between the current study and those of the Princeton trilogy. As a result, we conclude that (a) the results from Study 2 differed from the results reported in the Princeton trilogy because the stereotypes really have changed over time and (b) Study 1 failed to detect these changes because it used an outdated attribute list.

#### *THE CHANGING OF ETHNIC AND NATIONAL STEREOTYPES*

The results of Study 2 indicated that the ethnic and national stereotypes under investigation changed considerably in both content and consensus. The changes that emerged in consensus most often reflected increases rather than decreases. These results suggest that members of stereotyped groups may indeed confront different stereotypes than did past generations. However, because many of the new stereotypes are highly consensual, members of these groups may confront those stereotypes as often, and in some cases more often, than past generations confronted the stereotypes of their time.

#### *STUDY 3: THE FAVORABLENESS OF ETHNIC AND NATIONAL STEREOTYPES*

The social consequences of highly consensual stereotypes may partly depend on their favorableness. If stereotypes have become more favorable over time, then the potentially damaging effects that they have on members of stereotyped groups may be mitigated. In contrast, if stereotypes have become less favorable over time, then the benefits often associated with stereotype change may not materialize. People may repeatedly confront stereotypes that, although new, are just as damaging as old stereotypes. Therefore, Study 3 examined whether the ste-

reotypes have changed in favorableness over the past 60 years.

### Method

#### PARTICIPANTS

Participants were 19 White undergraduates (10 women and 9 men) enrolled in introductory psychology courses.

#### PROCEDURES AND QUESTIONNAIRE

Participants completed a questionnaire that assessed how favorably they perceived the attributes that have been endorsed as descriptive of the 10 ethnic and national groups at different points in time (i.e., 1930s, 1950s, 1960s, 1990s). The questionnaire contained 10 sections, one for each group under investigation. Each section first presented the group label, followed by a list of attributes. The list of attributes contained all the attributes endorsed as stereotypic of the group by each of the Princeton trilogy studies and those endorsed as descriptive of the group by the European American participants in Study 2 of this research. Thus, the questionnaire presented participants with all of the attributes comprising the 10 stereotypes at the four time periods (i.e., 1930s, 1950s, 1960s, 1990s). The attributes were presented to participants in random order. Participants rated the favorableness of each attribute in relation to the group it described using a 5-point scale with endpoints of  $-2$  (*very unfavorable*) and  $+2$  (*very favorable*).

### Results and Discussion

To determine the extent to which stereotypes have changed in favorableness over time, it would have been optimal to compare the favorableness ratings of the current participants to those of the Princeton trilogy participants. Unfortunately, only one Princeton trilogy study examined stereotype favorableness. Karlins et al.'s (1969) study investigated changes in favorableness by comparing how favorably their sample perceived the content of their own stereotypes to how favorably their sample perceived the content of the stereotypes reported by Katz and Braly (1933) and Gilbert (1951).

#### ASSESSMENT OF STEREOTYPE FAVORABLENESS

Borrowing Karlins et al.'s (1969) procedure, analyses assessed how favorably our sample perceived the contents of the stereotypes endorsed by each of the Princeton trilogy samples, as well as how favorably they perceived the contents of stereotypes endorsed by the European American sample in Study 2 of this research. For each group and for each participant, we calculated four favorableness scores by averaging the favorableness ratings for the four sets of attributes that comprised the

content of the stereotype at the four points in time—the 1930s, 1950s, 1960s, and 1990s.<sup>5</sup> For example, Gilbert (1951) reported that in the 1950s, Turks were perceived as cruel, very religious, treacherous, sensual, ignorant, physically dirty, and sly. Participants in Study 3 rated the favorableness of each of these seven attributes. We averaged each participant's ratings across these seven attributes to obtain one favorableness score per participant for the 1950s Turkish stereotype. We repeated this procedure three more times for the Turkish stereotype: once using the attributes reported by Katz and Braly (1933), a second time using the attributes reported by Karlins et al. (1969), and a third time using the attributes endorsed by the European American sample in Study 2 of this research. This procedure yielded four favorableness scores per participant for the Turkish stereotype, one pertaining to each of the four time periods (i.e., 1930s, 1950s, 1960s, 1990s). We performed these procedures for all 10 groups, creating a total of 40 favorableness scores per participant (i.e., 10 groups  $\times$  4 time periods per group).

#### CHANGES IN STEREOTYPE FAVORABLENESS

For each of the 10 groups, we performed a  $1 \times 4$  repeated-measures ANOVA. The four levels of the within-subject factor corresponded to the four time periods at which the content of each group's stereotype had been assessed (1930s vs. 1950s vs. 1960s vs. 1990s). The dependent variables in each analysis were the perceived favorableness of the stereotype at each of the four time periods. For all 10 groups, results showed that the stereotypes associated with the four time periods differed from each other with respect to favorableness (all  $F_s \geq 2.85$ ,  $p_s < .05$ ). Contrasts tested whether the favorableness of the stereotypes endorsed in the 1990s differed from the favorableness of the stereotypes endorsed in the 1930s, 1950s, and 1960s. Bonferroni corrections controlled for Type I error arising from these three comparisons by reducing the significance level to  $p \leq .017$ .

For the most part, results of the contrasts showed that participants considered stereotypes of the 1990s to be more favorable than past stereotypes. Participants perceived the Chinese, Japanese, and Turkish stereotypes of the 1990s significantly more favorable than the Chinese, Japanese, and Turkish stereotypes of the 1930s, 1950s, and 1960s (all  $t_s \geq 5.09$ ,  $p_s < .001$ ). Participants perceived the Irish stereotype of the 1990s significantly more favorable than the Irish stereotypes of the 1930s and 1960s (both  $t_s \geq 6.07$ ,  $p_s < .001$ ) and perceived the African American stereotype of the 1990s significantly more favorable than the African American stereotype of the 1930s and 1950s (both  $t_s \geq 4.97$ ,  $p_s < .001$ ). Participants perceived the Jewish stereotype of the 1990s as signifi-

cantly more favorable than the Jewish stereotype of the 1950s ( $t = 2.47, p = .017$ ) and perceived the Italian stereotype of the 1990s significantly more favorable than the Italian stereotype of the 1930s ( $t = 12.96, p < .001$ ). For only three groups did participants view past stereotypes as more favorable than contemporary stereotypes. Participants perceived the American stereotype of the 1990s significantly less favorable than the American stereotype of the 1930s and 1950s (both  $t_s \geq 3.67, p_s < .001$ ) and perceived the German and English stereotypes of the 1990s significantly less favorable than the German and English stereotypes of the 1930s, 1950s, and 1960s (all  $t_s \geq 3.51, p_s < .001$ ). There were no other significant differences in stereotype favorableness (all  $t_s \leq 2.44, p_s \geq .02$ ).<sup>6</sup>

The above results suggest that many of the stereotypes under investigation have become more favorable over time. This pattern seems to paint an encouraging picture. Perhaps the potentially damaging effects attributed to stereotypes have lessened over time. However, it is important to point out that even stereotypes that have become more favorable over time may still be unfavorable on an objective level. In addition, not all increases in stereotype favorableness have been steady and continuous. For example, the favorableness of the African American stereotype in the 1990s did not differ significantly from its favorableness in the 1960s ( $t = .91, p = .34$ ). This suggests that increases in stereotype favorableness may sometimes occur very slowly, especially among social groups that most often experience prejudice and discrimination.

#### GENERAL DISCUSSION

This article examined whether ethnic and national stereotypes have changed in content, consensus, and favorableness throughout the past 60 years by comparing current stereotypes of 10 ethnic and national groups to the stereotypes reported by the Princeton trilogy. Results indicated that the stereotypes underwent substantial revision over time. Almost all of the stereotypes changed in content and many changed in consensus. Most of the changes in consensus reflected increases rather than decreases, suggesting that current members of stereotyped groups may confront stereotypes more frequently than did past generations. However, the potentially damaging effects that highly consensual stereotypes can have on members of these groups may be partially tempered by the finding that most of the stereotypes became more favorable.

#### *Stereotype Change*

There exist several potential mechanisms that could result in stereotype change. Stereotypes can arise from

the roles that social groups occupy in society (e.g., Eagly & Steffen, 1984) and from the structure of relationships between social groups (e.g., Sherif, 1966). Change along these dimensions may alter the content of stereotypes. The nature of structural relationships may explain why European and non-European American participants in the current research sometimes endorsed different attributes when indicating their beliefs about the ethnic and national groups. There are important historical and contemporary differences in the relationships that these groups have with many of the ethnic and national groups that we examined. For example, European Americans are a majority in the United States and have economic and academic advantages in comparison to many non-European American groups. These advantages may buffer European Americans from unfair treatment, perhaps explaining why they perceived Americans as less prejudiced than did non-European Americans (see Table 1).

Stereotypes also may have changed because of increased intergroup contact (Allport, 1954). In the United States, people of different ethnicities and nationalities have more contact today than they did in previous decades (Dovidio & Gaertner, 1986). Indeed, segregation was legal when Katz and Braly (1933) and Gilbert (1951) performed their research. Increased contact may have given people the opportunity to revise the content of their stereotypes, thus contributing to stereotype change. This line of reasoning also suggests that ethnic and national stereotypes should have become less consensual over time. The more contact people have with members of a social group, the more heterogeneously they tend to perceive that social group (e.g., Linville, Fischer, & Salovey, 1989). However, our results indicated that ethnic and national stereotypes tended to become more consensual, despite the possibly greater contact between members of different social groups (see Lee & Ottati, 1993, for related findings). These results suggest that stereotype change is a complex process that may be affected by a variety of factors in addition to contact.

One factor that might have played a role in stereotype change is perceptual accuracy. Stereotypes sometimes reflect the actual characteristics of groups (e.g., Judd & Park, 1993; Lee, Jussim, & McCauley, 1995; Lee & Ottati, 1993). For example, teachers accurately perceive differences and similarities between students of different sex, social class, and ethnic groups (Madon et al., 1998). If stereotypes are sometimes accurate, then they must also sometimes change when the social groups change. Thus, the indications of stereotype change suggested by the current research may partly reflect changes that the stereotyped groups have actually undergone.

*Barriers to Fair Treatment  
Despite Changing Stereotypes*

Stereotype change does not, however, ensure fair treatment. Even generally favorable stereotypes may include unfavorable subtypes (Deaux, Winton, Crowley, & Lewis, 1985). Favorableness also may be context dependent. For example, although women are generally perceived more favorably than men, they may experience job discrimination because the positive qualities that they supposedly possess are devalued in high-paying professions (Eagly, Mladinic, & Otto, 1991). Barriers to equal opportunities also may exist in the absence of bias associated with stereotypes. For instance, employment opportunities are often communicated through informal networks that may not reach across the boundaries of ethnic and national groups (Braddock & McPartland, 1987). Because European Americans are more often connected to such networks, they will tend to have greater job opportunities than will other ethnicities. Finally, mere knowledge of a cultural stereotype may produce bias among high-prejudice people or among perceivers in general when only an individual's social group membership is known (Devine, 1989; Lepore & Brown, 1997).

*Limitations*

*Content over process.* This research focused on issues related to stereotype content, as opposed to issues of stereotype process. Stereotype content studies may not generalize to stereotypes not under investigation. For example, we focused on groups whom Katz and Braly (1933) believed were salient during the 1930s. Had we assessed changes in stereotype favorableness of currently salient groups (e.g., Serbs, Iraqis, Mexicans), our results may have revealed a trend toward decreased favorableness. This limitation notwithstanding, stereotype content studies lay the groundwork for examining the processes by which stereotypes may create social problems. Moreover, all reviews of stereotypes in the past 50 years define stereotypes as involving the content of people's beliefs about social groups (e.g., Allport, 1954; Ashmore & Del Boca, 1981; Brigham, 1971; Lee et al., 1995). Identifying the content of stereotypes, therefore, is of fundamental importance for acquiring a more complete understanding of stereotypes.

*Explicit versus implicit assessment measures.* This research employed explicit measures to assess the stereotypes. Explicit measures may not correspond to implicit measures of stereotypes (e.g., Devine, 1989), particularly when assessing stereotypes of groups about whom it is socially undesirable to report unfavorable perceptions.

This raises the possibility that in an attempt to appear unbiased our participants were more reluctant to endorse unfavorable stereotypes than were participants in the Princeton trilogy studies. However, the pattern of current findings suggests that social desirability concerns did not have a major impact on participants' responses. First, participants were aware that their responses were anonymous. Second, if social desirability had been a major factor, then participants would have shied away from endorsing unfavorable stereotypes. However, the stereotypes reported in Study 1 were highly similar to the stereotypes reported by the Princeton trilogy, despite the fact that some were quite unfavorable. Although the stereotypes reported in Study 2 were different and more favorable than those reported by the Princeton trilogy, Study 2 used an updated attribute list. Thus, participants in the current research readily endorsed unfavorable stereotypes when offered only Katz and Braly's (1933) attribute list in Study 1 but endorsed more favorable stereotypes when offered the updated attribute list in Study 2. Finally, analyses of the favorableness data were within-subject comparisons of the stereotypes as determined at four points in time. This design controls for between-subject differences in social desirability across the four time periods with respect to any particular group, especially because attributes were presented in random order so that participants could not identify which attributes were from the 1930s, 1950s, 1960s, and 1990s.

CONCLUSION

Stereotypes have traditionally been viewed as resistant to change, highly consensual, and unfavorable (see Lee et al., 1995, for a review). These assumptions imply that stereotype change is difficult to achieve and that even if stereotypes do change, that change may not translate into improved intergroup relations because new stereotypes may be just as unfavorable and consensual as the stereotypes that they replaced. Our research addressed these issues by examining the extent to which 10 ethnic and national stereotypes have changed in content, consensus, and favorableness across more than 60 years. Our results showed that many of the stereotypes are still highly consensual but that they have changed substantially in content and that most became more favorable. Although these changes may have contributed to improvements that have occurred in intergroup relations within the United States during the past several decades, the relationship between stereotype change and intergroup relations is probably characterized by reciprocal effects. Thus, the stereotype change reported

in the current research may be both a cause and a consequence of the improvement in relations between ethnic and national groups that have occurred in the United States across the 20th century.

#### NOTES

1. Stereotype change was not examined with the non-European American sample because the Princeton trilogy relied exclusively on White participants.

2. Nonparametric statistics were used because the percentages were not normally distributed.

3. We used Gough and Heilbrun's adjective checklist because it is among the most comprehensive checklists available, including an extensive collection of descriptive adjectives that cover a broad range of personality traits.

4. Forty-five traits on Gough and Heilbrun's (1983) adjective checklist are identical to those used by the Princeton trilogy and 1 trait (i.e., pleasure seeking) is very similar to a trait used by the Princeton trilogy (i.e., pleasure loving). This left 254 unique traits from Gough and Heilbrun's adjective checklist that were included in the updated attribute list.

5. Stereotype favorableness was based on all attributes in a stereotype, including ties. For example, the favorableness of the American stereotype (Study 2: European American sample) equaled the average favorableness scores of the 19 attributes in that stereotype (see Table 1).

6. A sample of non-European Americans ( $n = 14$ ; 6 African Americans, 1 Indian, 5 Asians, and 2 Latino/as) was surveyed to examine differences in stereotype favorableness between European and non-European Americans. A questionnaire presented participants with the 10 group labels and the attributes that were endorsed as descriptive of each group by previous participants (i.e., participants in the Princeton trilogy and the non-European American participants in Study 2 of this article). The results from 10 separate independent sample  $t$  tests (1 for each of the 10 groups) indicated that the non-European American sample endorsed significantly more favorable stereotypes about Germans and Turks (both  $t_s \geq 2.27$ ,  $p_s \leq .03$ ) and significantly less favorable stereotypes about Americans and Italians (both  $t_s \geq 2.93$ ,  $p_s \leq .01$ ) than did the European American sample from Study 3 of this research. There were no other significant differences between the European and non-European American samples (all  $t_s \leq 1.72$ , all  $p_s \geq .10$ ).

#### REFERENCES

- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Amicus Curiae Brief for the American Psychological Association. (1991). In the Supreme Court of the United States: *Price Waterhouse v. Ann B. Hopkins*. *American Psychologist*, 46, 1061-1070.
- Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 37-81). Hillsdale, NJ: Lawrence Erlbaum.
- Ashton, M. C., & Esses, V. M. (1999). Stereotype accuracy: Estimating the academic performance of ethnic groups. *Personality and Social Psychology Bulletin*, 25, 225-236.
- Braddock, J. H., & McPartland, J. M. (1987). How minorities continue to be excluded from equal employment opportunities: Research on labor market and institutional barriers. *Journal of Social Issues*, 43, 5-40.
- Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, 76, 15-33.
- Bruner, J. S. (1957). Going beyond the information given. In H. E. Gruber, K. R. Hammond, & R. Jessor (Eds.), *Contemporary approaches to cognition* (pp. 41-69). Cambridge, MA: Harvard University Press.
- 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.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Deaux, K., Winton, W., Crowley, M., & Lewis, L. L. (1985). Level of categorization and content of gender stereotypes. *Social Cognition*, 3, 145-167.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5-18.
- Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin*, 21, 1139-1150.
- Dovidio, J. F., & Gaertner, S. L. (Eds.). (1986). Prejudice, discrimination and racism: Historical trends and contemporary approaches. In *Prejudice, discrimination and racism* (pp. 1-30). Orlando, FL: Academic Press.
- Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? *Psychology of Women Quarterly*, 15, 113-137.
- Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, 46, 735-754.
- Fiske, S. T., & Glick, P. (1995). Ambivalence and stereotypes cause sexual harassment: A theory with implications for organizational change. *Journal of Social Issues*, 51, 97-115.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances In Experimental Psychology*, 23, 1-74.
- Gibbons, J. D., & Chakraborti, S. (1992). *Nonparametric statistical inference* (3rd ed.). New York: Marcel Dekker.
- Gilbert, G. M. (1951). Stereotype persistence and change among college students. *Journal of Abnormal and Social Psychology*, 46, 245-254.
- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70, 491-512.
- Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychology of Women Quarterly*, 21, 119-135.
- Gough, H. G., & Heilbrun, A. B. (1983). In O. K. Buros (Ed.), *The Adjective Checklist Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100, 109-128.
- Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 281-388). San Diego, CA: Academic Press.
- Karlins, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 13, 1-16.
- Katz, D., & Braly, K. W. (1933). Racial stereotypes of one-hundred college students. *Journal of Abnormal and Social Psychology*, 28, 280-290.
- Kunda, Z., Sinclair, L., & Griffin, D. (1997). Equal ratings but separate meanings: Stereotypes and the construal of traits. *Journal of Personality and Social Psychology*, 72, 720-734.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- LaPierre, R. T. (1936). Social forces type-rationalizations of group antipathy. *Social Forces*, 15, 232-237.
- Lee, Y. H., Jussim, J., & McCauley, C., R. (Eds.). (1995). *Stereotype accuracy: Toward appreciating group differences*. Washington, DC: American Psychological Association.
- Lee, Y. H., & Ottati, V. (1993). Determinants of in-group and out-group perceptions of heterogeneity: An investigation of Sino-American stereotypes. *Journal Of Cross-Cultural Psychology*, 24, 298-318.

- Lepore, L., & Brown, R. (1997). Category and stereotype activation. Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*, 275-287.
- Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, *57*, 165-188.
- Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin*, *24*, 1304-1318.
- McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, *36*, 929-939.
- Seago, D. W. (1947). Stereotypes: Before Pearl Harbor and after. *Journal of Psychology*, *23*, 55-64.
- Sherif, M. (1966). *Group conflict and cooperation*. London: Routledge Kegan Paul.
- Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, *18*, 247-255.
- Stangor, C., & Lange, J. (1994). Mental representations of social groups: Advances in understanding stereotypes and stereotyping. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 357-416). San Diego, CA: Academic Press.

Received October 27, 1999

Revision accepted August 7, 2000