

---

# Perceptions of Trait Typicality in Gender Stereotypes: Examining the Role of Attribution and Categorization Processes

Joachim I. Krueger

Julie F. Hasman

Melissa Acevedo

Brown University

Paola Villano

Università di Bologna

---

*Gender stereotypes are understood as the ascription of different personality traits to men and women. Data from American and Italian samples showed that consistent with the attribution hypothesis, the estimated prevalence of a trait in a target group predicted perceptions of trait typicality well. In contrast, there was no support for the categorization hypothesis, according to which perceived differences in trait prevalence between groups should independently predict trait typicality. Nevertheless, participants overestimated gender differences in personality as predicted by the principle of intercategory accentuation. The implications of these findings for the rationality and accuracy of gender stereotyping are discussed.*

---

In the area of stereotype research, the study of gender stereotypes has held a prominent place (Diekmann & Eagly, 2000). Among the reasons for this prominence are the pervasiveness of these stereotypes, their linkages to gender roles, and the simple fact that nearly anyone who stereotypes is also a member of one of the two gender groups. As is true for most social stereotypes, gender stereotypes comprise a variety of personality-descriptive trait terms seen as typical or stereotypical of men or women. There is little doubt that gender stereotypes exist or that most people know what they are, but it is less clear just how perceptions of trait typicality arise. The perceived typicality of attributes seems to be a forgotten issue in stereotype research, where more studies are concerned with the typicality of instances (i.e., group members) than with the typicality of their features (e.g., Rothbart, Sriram, & Davis-Stitt, 1996). Still, the record of the field suggests that there are two hypotheses regard-

ing this question, and the goal of the present research is to test them simultaneously in a competitive manner.

The attribution hypothesis assumes a simple associationist process by which people learn and encode the properties of social groups. When forming mental associations between traits and groups, people attend to the prevalence of various traits in a target group. When a trait has been frequently and recently paired with a group, people come to see it as prevalent in the group and judge its typicality accordingly. Early behaviorist work favored this hypothesis. Preferences for nations, for example, can be classically conditioned by pairing positive versus negative words with the group labels (Staats & Staats, 1958). Such associations can even be formed outside of awareness (Olson & Fazio, 2001).

The categorization hypothesis assumes a more complex process of comparative reasoning. People are assumed to learn and encode differences between salient groups and to judge the typicality of each property depending on these differences. The advantage of being sensitive to group differences is that it permits the categorization of a target individual on the basis of trait information. When an individual exhibits a trait that is

---

**Authors' Note:** We thank Jason Taylor for help with S-PLUS and Ursula Athenstaedt and Robyn Dawes for comments on a draft version of this article. We are also indebted to Larry Barsalou and Rob Goldstone for educating us on the cognitive psychological literature on typicality effects. Correspondence should be addressed to Joachim Krueger, Department of Psychology, Brown University, Box 1853, 89 Waterman St., Providence, RI 02912; e-mail: joachim\_krueger@brown.edu.

*PSPB*, Vol. 29 No. 1, January 2003 108-116

DOI: 10.1177/0146167202238376

© 2003 by the Society for Personality and Social Psychology, Inc.

perceived to be more common in the target group than in a comparison group, the perceiver can categorize him or her as a group member. The drawback of this process is that comparative reasoning requires greater mental resources and discipline than the automatic, associationist reasoning implied by the attribution hypothesis. Historically, the categorization hypothesis has held great appeal for theories concerned with the interplay between the self and social groups. In particular, theories of social identity and self-categorization consider the contextual nature of the perception of groups essential (e.g., Oakes, Haslam, & Turner, 1994). According to these theories, a sense of belonging to a group—and thus a sense of social identity—can only arise to the extent that the group can be differentiated from relevant other groups.

Perhaps because the attribution and the categorization hypotheses arose from separate intellectual traditions, comparative tests have been rare, and there have been none in the area of gender stereotyping. This is surprising because both hypotheses have long been known as distinctive perspectives on stereotyping. Zawadski (1948) wrote that

the popular conception of a group characteristic seems to be: "a characteristic which is present in the majority of the members of the group." According to this concept, it is a necessary and sufficient condition for a group characteristic to be represented in at least 51 per cent of the members of the group. A moment's reflection, however, suggests another concept. A group characteristic which makes possible a distinction between two groups. Such a distinction is possible, for instance, if the group A possesses the characteristic *c* in 40 per cent and the group B in 20 per cent. (pp. 135-136)

Which of these conceptions describes more accurately what people actually do, Zawadski could not say. And so, perhaps for simplicity, most empirical work on stereotype assessment proceeded as if the attribution hypothesis were true (Brigham, 1971). Meanwhile, some theorists continued to caution, as Zawadski had done, that stereotypes may involve an element of categorization. Among several stereotype indexes, Allport (1954) included what he called "categorical differentials [which occur] when some single attribute is found with differential frequency in various groups" (p. 102). Going further, Tajfel (1969) identified the perceptual accentuation of group differences as the heart of stereotyping (see also Campbell, 1956). Accentuation theory was one of the precursors of self-categorization theory with its emphasis on "meta-contrasts" (e.g., Oakes et al., 1994). Self-categorization theory assumes that people stereotype themselves and others by minimizing perceived differences within social groups and by maximizing differ-

ences between groups. Self-categorization theory owes some of its ideas to Rosch's prototype theory of categorization (e.g., Rosch & Mervis, 1975). The perceived prototypicality of instances, according to Rosch, depends jointly on their similarity with other category members and their dissimilarity with nonmembers. Unlike their social psychological counterparts, however, prototype theories addressing the categorization of natural kinds put little stock in the possibility of perceptual distortions. Perhaps because of this sustained theoretical activity, the categorization hypothesis has begun to dominate empirical work on stereotype measurement (Judd & Park, 1993), experiments concerned with stereotype acquisition (Ford & Stangor, 1992), and models of stereotype representation (Kunda & Thagard, 1996).

In the empirical literature, the seminal article by McCauley and Stitt (1978) marked the shift away from the attribution hypothesis. McCauley and Stitt conceptualized—and defined—stereotypes as perceived differences between group characteristics and trait base rates in the world. Assuming that stereotypes are "composed of those attributes for which within-group predictions differ from base-rate predictions" (p. 938), McCauley and Stitt asked American college students to estimate the percentages of Germans possessing various traits and to estimate the corresponding percentages for the world population. For each participant, they correlated Likert-type scale typicality ratings with the diagnostic ratios of estimates for Germans over estimated base rates. On average, these idiographic correlations were substantial ( $M = .58$ ). The case for the categorization hypothesis was not conclusive, however, because the estimates for the Germans alone (i.e., the ratio's numerator) predicted typicality ratings even better ( $M = .74$ ). Thus, the division by base rates actually reduced the predictability of typicality ratings. A later study on national stereotypes showed that the diagnostic ratio contributed nothing to the prediction of trait typicality ratings once the simple percentage estimates for the target group were controlled (Krueger, 1996b).

Although such results cast some doubt on the empirical utility of the categorization hypothesis, it is not clear whether they can be generalized to gender stereotypes. Because there are only two mutually exclusive and exhaustive gender groups, their stereotyped descriptions may be more dependent on intercategory comparisons and contrasts. Following McCauley and Stitt (1978), Martin (1987) collected percentage estimates from men and women about men and women on a set of trait adjectives that were known to be masculine, feminine, or neutral (Bem, 1981; Appendix A). Martin's main interest was to show that for most items, the diagnostic ratios computed from perceived percentages were more extreme than diagnostic ratios computed from the self-

descriptions of men and women in a criterion sample. Reanalysis of her data (from Table 3, p. 493) showed that the average diagnostic ratio was 1.4 for the masculine traits and 0.7 for feminine traits. This result suggests that masculine traits were stereotypic of men because they were perceived to be more common among men than among women. Similarly, feminine traits appeared to be stereotypic of women because they were perceived to be less common among men than among women. Note, however, that the same data also are consistent with the attribution hypothesis. For each gender, a diagnostic ratio could be computed by dividing the estimates for masculine traits by the estimates for feminine traits. This method captured perceived differences between traits instead of differences between groups. The ratio was 1.3 for men and 0.7 for women.

Neither McCauley and Stitt (1978) nor Martin (1987) addressed the natural confound between trait attribution and social categorization. According to the attribution hypothesis, the perceived typicality of a trait increases with the probability that the trait is present given that the individual is, say, a man (i.e.,  $p[T|M]$ ). According to the categorization hypothesis, the perceived typicality of a trait increases with a ratio of two probabilities (e.g., the probability of the trait assuming that the individual is a man divided by the probability of the trait assuming that the individual is a woman,  $p[T|M]/p[T|W]$ ).<sup>1</sup> Because trait attributions to the target group are in the numerator, the diagnostic ratio increases with increases in trait attributions to the target group. The more prevalent a trait is thought to be in the target group, the more likely it is, by regression to the mean alone, that the same trait is thought to be less prevalent in the comparison group. Using a computer simulation, we found that simple trait attributions for a target group are highly confounded with both ratios and difference scores (Appendix B). To capture the unique contribution of trait attributions to the prediction of trait typicality, it is necessary to hold the diagnostic ratios constant; to capture the unique contribution of attribution to the prediction of trait typicality, it is necessary to hold trait attributions to the target group constant.

Some of the underlying theoretical work on categorization (e.g., accentuation theory, self-categorization theory) and some of the empirical work (Martin, 1987, but not McCauley & Stitt, 1978) suggests that only the categorization hypothesis is consistent with the idea that people often overestimate group differences. We remained agnostic on this issue. Indeed, we could imagine that people derive percentage estimates for a target group in part by contrasting them away from corresponding estimates in the comparison group (Krueger, 1992; Spears, Doosje, & Ellemers, 1997). If so, simple trait attributions

to a target group already imply intergroup contrasts, leaving little to explain to the ratios of the estimates.

Following McCauley and Stitt (1978) and Martin (1987), our analysis focuses on ascribed attributes (i.e., trait adjectives) and not on identifying group attributes. The reason is that nearly everyone in the target group but hardly anyone in the opposite group shares gender-identifying attributes, such as sexual characteristics. Thus, any tests of the attribution hypothesis relative to the categorization hypothesis would be totally confounded.

#### *Method*

We recruited two samples. The American sample comprised 42 male and 44 female undergraduate students at Brown University; the Italian sample comprised 42 male and 42 female students at the Università di Bologna. Assuming that the core of gender stereotypes is "pancultural" (Williams, Satterwhite, & Best, 1999), we expected converging results. Age information was not gathered for the American sample, but judging from the typical participant pool, we estimated the median age to be 19 years. The median age in the Italian sample was 22 years. Both samples were recruited in various locations around campus, and participants were run individually.

Similar to Martin (1987), we used 10 masculine traits, 10 feminine traits, and 10 gender-neutral traits from the Bem Sex Role Inventory (BSRI) (Bem, 1981) as stimulus materials (Appendix A). One of the authors (PV) translated the items into Italian and a bilingual graduate student translated the Italian list back to English. Two of the authors (JK, PV) discussed and removed the few remaining discrepancies. Participants rated the 30 traits in alphabetical order. Different instructional sets were presented on separate sheets. Participants made three ratings for each target sex: (a) an estimate of the percentage of men (or women) who could be described by the trait, (b) a rating of how typical the trait was of the target sex in the participant's personal view, and (c) a rating of how typical the trait was of the target sex in terms of the cultural stereotype. Percentage estimates were meant to reflect "your own personal impression." Typicality ratings could range from 1 (*not typical at all*) to 9 (*very typical*). Instructions for the cultural stereotype ratings stressed that "ratings may or may not coincide with your own personal views. What we are interested in are your perceptions of the cultural stereotypes." Each possible order of the three sets of ratings (percentage estimates, typicality ratings, cultural stereotypes) was presented to some participants, and the order of the two gender target groups was varied within each set of ratings.

#### *Results and Discussion*

*Preliminaries.* Because of the complexity of the data and the variety of available analytical options, two

prefatory notes are in order. First, exploratory analyses revealed no meaningful effects involving the gender of the target group or the gender of the participants (all  $p > .10$ ; see also Beyer, 1999; Conway & Vartanian, 2000; or Hall & Carter, 1999, for the lack of the latter). For economy of presentation, we ignored these variables. Wishing to emphasize the cross-cultural generality of the findings, however, we tested the substantive hypotheses separately for the American and the Italian samples. Second, we emphasized idiographic analyses by computing correlations for each participant and across traits. Each correlation was transformed to its corresponding Fisher  $Z$  score, whose averages were transformed back to correlation coefficients. Because standard errors were small (ranging between .03 and .06), we limited tests of statistical significance to those effects whose significance was not self-evident. To demonstrate the robustness of the findings, we note their convergence with nomothetic analyses, in which correlations are computed for individual traits and across participants, and we replicate the analyses using ratings of cultural stereotypes as the criterion variables.

*Instrument validity.* The BSRI continues to be widely used as a measure of perceived sex roles. Most recent efforts at validation have confirmed Bem's (1981) original psychometric work (Auster & Ohm, 2000; Holt & Ellis, 1998; but see Wilcox & Francis, 1997). Confirming that masculine and feminine traits are still primarily ascribed to their respective genders, we found in the American sample that the mean typicality ratings for masculine traits were 6.63 for men and 5.15 for women; 4.53 and 6.60, respectively, for feminine traits; and 5.24 and 6.06 for neutral traits. The differences for all 20 gender-related traits were significant in univariate analyses. The findings in the Italian sample were much the same,  $M_{\text{masculine traits}} = 6.48$  (men) and 5.49 (women);  $M_{\text{feminine traits}} = 5.10$  (men) and 6.83 (women);  $M_{\text{neutral traits}} = 5.48$  (men) and 6.43 (women). Only the masculine traits of "defends beliefs" and "independent" failed to yield significant differences.

*Attribution versus categorization.* How well do percentage estimates (attribution) and the diagnostic ratios (categorization) predict ratings of trait typicality? We computed modified diagnostic ratios to limit their range from  $-1$  to  $+1$  and to ensure that the confound between attribution and categorization was expressed by positive correlations with typicality ratings for both target groups. Whichever percentage estimate was smaller (for men or for women) was the numerator. If the estimate for the target group was in the numerator, the value of 1 was subtracted from the ratio. If the estimate for the comparison group was in the numerator, the ratio was subtracted from 1. In both samples, modified ratios,

conventional ratios, and difference scores were correlated almost perfectly across traits and within participants ( $M > .93$ ).

Figure 1 shows the mean idiographic zero-order correlations and the mean partial correlations (in brackets). Because every participant judged both target sexes, each mean was based on 172 correlations in the American sample and on 168 correlations in the Italian sample. Not surprisingly, trait attributions were highly correlated with their corresponding diagnostic ratios, and both predicted typicality ratings. The partial correlations indicated that the attribution hypothesis accounted for the data better than did the categorization hypothesis. In contrast, the partial correlations between typicality ratings and diagnostic ratios reached significance only in the American sample,  $t(171) = 6.46$ . Across participants, we found a hydraulic relationship between reliance on trait attributions and reliance on the intergroup comparisons (i.e., the diagnostic ratio). As the unique predictive power of trait attributions increased across respondents, the unique predictive power of the diagnostic ratios decreased ( $r_{\text{American}} = -.50$ ,  $r_{\text{Italian}} = -.39$ ).

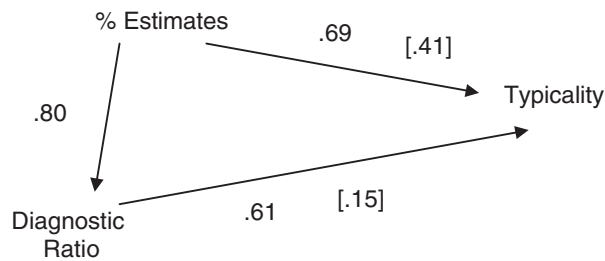
To ensure that the tests of the attribution and the categorization hypotheses were not limited to our specific methodological choices, we conducted four follow-up analyses. First, we computed nomothetic analyses by treating each of the 30 traits (as opposed to each participant) as the unit of analysis. As expected on mathematical grounds (Kenny & Winqvist, 2001), the same findings emerged. The partial correlations between typicality ratings and percentage estimates for the target group were of considerable magnitude ( $M_{\text{American}} = .53$ ,  $M_{\text{Italian}} = .48$ ), whereas the partial correlations between typicality ratings and diagnostic ratios were essentially zero ( $M_{\text{American}} = .04$ ,  $M_{\text{Italian}} = -.006$ ).

Second, we returned to the idiographic analyses and compared the correlations between typicality ratings and percentage estimates for the target gender ( $M_{\text{American}} = .69$ ,  $M_{\text{Italian}} = .60$ ) with the correlations between typicality and estimates for the opposite gender ( $M_{\text{American}} = -.27$ ,  $M_{\text{Italian}} = -.09$ ). The finding that the former correlations were much larger in absolute terms than the latter helps explain the low predictive power of the diagnostic ratio.

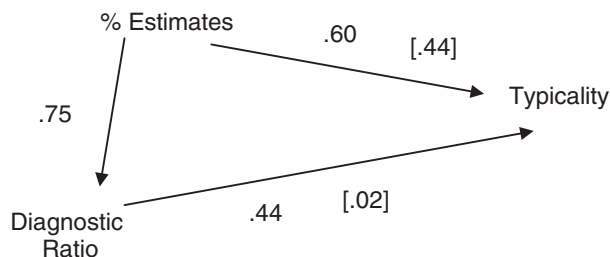
Third, we performed the idiographic analyses separately for each of set of gender-related traits (masculine, feminine, or neutral) before aggregating these results. As in the omnibus analysis, the unique attribution effects were robust ( $M_{\text{American}} = .36$ ,  $M_{\text{Italian}} = .42$ ), whereas the unique categorization effects disappeared ( $M_{\text{American}} = .00$ ,  $M_{\text{Italian}} = -.02$ ).

Fourth, we used ratings of cultural stereotypes instead of personal typicality ratings as criterion variables. In an analysis of idiographic correlations across all 30 traits, our initial findings replicated in the Italian sample.

## AMERICAN Sample



## ITALIAN Sample



**Figure 1** Perceptions of how typical a trait is of a gender group as predicted by trait attributions (i.e., percentage estimates) and diagnostic ratios.

NOTE: Mean partial correlations are in brackets.

Here, the unique contribution of percentage estimates for the target group ( $M = .28$ ) was greater than the unique contribution of the diagnostic ratio ( $M = .11$ ),  $t(167) = 3.76, p < .001$ . In the American sample, however, the predictor variables were roughly equivalent ( $M = .25$  and  $.30$ ),  $t = 1.02$ . One is tempted to conclude that Americans but not Italians think comparatively when describing the genders in terms of cultural stereotypes. The reasons for this discrepancy remain unclear and await further study. Similar to others before us (McCauley & Stitt, 1978), we considered personal typicality ratings as the primary testing ground for the attribution and the categorization hypotheses because percentage estimates also were generated from the personal perspective. Although ratings of cultural stereotypes were closely related to personal typicality ratings ( $M_{\text{American}} = .62, M_{\text{Italian}} = .54$ ),<sup>2</sup> they involve, at least in part, people's notions of what other people might think, and thus perhaps a somewhat less associative mental process.

*Accentuation.* The modest predictive value of the diagnostic ratio does not, of course, imply that people do not

accentuate gender differences. We assumed that (exaggerated) perceptions of intergroup differences are already encoded in the percentage estimates for the target groups. In the first analysis, the idiographic correlations between trait attributions for men and trait attributions for women served as measures of accentuation. Analogous correlations were computed for typicality ratings and for ratings of cultural stereotypes. The correlations across all 30 traits revealed that participants assumed dissimilarities between the genders. All six means were negative ( $M = -.24$  and  $-.10$  for percentage estimates in the American and the Italian sample, respectively;  $-.28$  and  $-.11$  for typicality ratings; and  $-.56$  and  $-.29$  for ratings of cultural stereotypes) and all differed significantly from zero ( $p < .05$ ). When correlations were computed within each category of gender-typed items, the averages were close to zero (max =  $.12$ , min =  $-.10$ ). Either way, these correlations were much smaller than the correlation between the average self-descriptions for men and women in Martin's (1987) criterion sample (i.e.,  $r = .85$ ; see Table 3, p. 493).

Psychometrically, these low correlations may not only indicate a lack of people's awareness that men and women describe themselves similarly but also, at least in part, an artifact of unreliability. Individual participants were treated much like measurement instruments intended to detect actual gender similarities and differences. The reliability of individual instruments can be increased through aggregation. As suggested by the Spearman-Brown correction, and as proved by Dawes (1970), correlations between composites are more extreme than composites of correlations. Thus, it should come as no surprise that correlations between average ratings across the 30 traits were far more negative than the averages of the idiographic correlations,  $r(\text{estimates}) = -.75$  and  $-.46$  in the American and the Italian sample, respectively;  $r(\text{typicality}) = -.77$  and  $-.53$ ;  $r(\text{cultural stereotypes}) = -.91$  and  $-.80$ . Because individuals have a slight tendency to see the genders as opposites, accentuation gains in strength when considered a group phenomenon.

#### Conclusions and Implications

As suggested by the attribution hypothesis, estimates of trait prevalence for a target group emerged as the primary predictor of trait typicality ratings. Comparative thinking, as captured by the diagnostic ratio, had little incremental value. This does not negate, however, the importance of categorization as part of stereotyped thinking, nor does it contradict the idea that people overestimate group differences. Indeed, raters in both samples accentuated gender differences; they simply did not use these perceived group differences to judge the typicality of the traits for any particular gender. Given

the repeated-measures design of the study, one may wonder if participants accentuated gender differences more than they would have if they had rated only one gender. This hypothesis appears to be false. A follow-up study using a single gender as a target group also supported the attribution hypothesis (Hasman & Krueger, 2002).

*A note on compound predictors (ratios and differences).* The contest between the attribution and the categorization hypotheses recalls earlier debates on the utility of difference scores as predictors of self-enhancement. In a notable exchange, Colvin, Block, and Funder (1995) argued that the differences between self-evaluations and peer evaluations crucially, and inversely, predict psychological adjustment. In response, Zuckerman and Knee (1996) argued that difference score measures (or ratios for that matter) cannot predict anything that is not already contained in the variables that constitute the differences. Ultimately, Asendorpf and Ostendorf (1998) supported Zuckerman and Knee's methodological point by way of an ingenious psychometric analysis but nevertheless endorsed Colvin et al.'s theoretical stance. Some compound measures are so appealing conceptually that they survive methodological challenges.

Examining self-enhancement with a different kind of difference score, Klar and Giladi (1999) studied how people judge their own happiness relative to the happiness of others. Consistent with a "self-focus model," only absolute ratings of one's own happiness predicted comparative ratings. Absolute ratings of the average person's happiness, and thus the differences between self- and other-ratings, were irrelevant. People simply thought they were happier than others inasmuch as they considered themselves to be happy. The parallels to the present research are obvious, although it remains to be seen which fate will befall the diagnostic ratio (and thus the categorization hypothesis) in stereotyping. After all, its theoretical appeal is considerable.

*Learning associations.* The theoretical underpinnings of the attribution hypothesis remain thinner than those of the categorization hypothesis. At the outset, we noted that simple learning processes such as classical conditioning are sufficient to establish trait-group associations. Because the occurrence of classical conditioning is usually inferred from a change in sentiment (e.g., greater liking for a nation), this account cannot fully explain how probabilistic trait attributions (percentage estimates) come about. We assume that trait attributions arise, in part, from observed or communicated frequency information. People can convert frequencies into probabilities (Estes, 1976), and when no other information is available, they use frequencies to form impressions of individual (Lambert, 1995; Sherman,

1996) or multiple categories (Fried & Holyoak, 1984; Krueger, 1992).

The goal of a recent experiment was to find out if, when given a choice, people seek frequency information that allows them to estimate probabilities relevant for trait attributions or probabilities relevant for categorization (Krueger, 2002). Participants were told about a set of 60 people. Each person was a member of one of two groups within the set (say A or B), and each was characterized by one of two traits (say X and Y). The only piece of frequency information given was that 18 people were members of Group A and possessed Trait X. Participants were charged with judging how typical Trait X was of Group A. To do that, they needed to select other frequency information. The great majority preferred to learn how many members of Group A had Trait Y rather than how many members of Group B had Trait X. In other words, participants were more concerned with estimating the probability of the trait within the group (i.e., the trait attribution) than with the probability that a person with the trait also was a member of the target group (i.e., categorization). This basic pattern held across a number of replications in which either the group labels were meaningful (e.g., men and women) and the traits were meaningless (X and Y) or in which the traits were meaningful (friendly and intelligent) and the group labels were not.

Of course, there would be no accentuation of group differences if frequency learning were the only determinant of group or category representation. Thus, several lines of research suggest that people forge a compromise between being sensitive to group averages, frequencies, and intergroup differences (Barsalou, 1985; Krueger, Rothbart, & Sriram, 1989). Applied to the context of gender stereotyping, this means that people represent men and women, in part, as idealizations of what is uniquely male or female.

*The rationality of stereotyping.* It is often assumed that stereotypes reflect irrational mental activity (Ashmore & Del Boca, 1981). Periodically, however, this assumption has been challenged (Allport, 1954; Tajfel, 1969). In the present context, one cannot judge whether the formation of trait-group associations is rational, but the exclusive use of these associations for making typicality ratings becomes an issue. Being associationist rather than comparative, trait attributions show the properties of nonrational mental heuristics that yield acceptable results most of the time while also producing systematic errors.

Depending on one's definition of rationality, one can focus on errors of inaccuracy or on errors of incoherence. By the first definition, it is important to understand the long-term consequences of a mental process. In ste-

reotyping, the accuracy of trait attributions can be assessed if there is an adequate reality criterion of what stereotyped groups are really like (e.g., locally computed performance measures, Beyer, 1999; meta-analytic effect sizes of group differences, Hall & Carter, 1999; or data from national probability samples, Diekmann, Eagly, & Kulesa, 2002). In research involving personality traits, the criterion question is often addressed by proxy variables, such as the similarity of autostereotypes and heterostereotypes or the similarity between stereotypes and the group members' self-descriptions. In Martin's (1987) data, for example, there was a modest correlation between average trait attributions and self-descriptions in the criterion sample ( $r = .23$ ; see Table 3, p. 493). Alternatively, accuracy could be assessed by the correlation between the diagnostic ratios derived from percentage estimates and the diagnostic ratios derived from criterion percentages. The low correlation obtained with this approach ( $r = .05$ ) may not be surprising because ratios involve two potential sources of error, whereas simple trait attributions involve only one.<sup>3</sup> This difference contradicts the idea that associationist judgments, because they are less rational, are also less accurate than judgments involving comparisons.

By the second definition of rationality, comparative but not associationist thinking preserves the internal consistency or coherence of various types of judgments. Judgments are deemed rational if they avoid outright contradictions. One way to appraise the coherence of probability estimates is to ask whether they conform to Bayes's Theorem (Dawes, 1998). Our data suggest that they do not. When judging the typicality of a trait simply by the perceived prevalence of the trait in the group, people ignore their own perceptions of how prevalent the trait is in the comparison group. This neglect leads to incoherence when people categorize individuals into groups based on the presence or the absence of certain traits. Failure to discriminate between judgments of attribution and judgments of categorization may result in the same person being categorized into mutually exclusive groups whenever the trait is seen as typical of both (for a related discussion of the irrationality of self-focus, see Krueger & Mueller, 2002).

We stress this discussion of two types of (ir)rationality because they can conflict with each other such that an increase in one bias entails a decrease in another. In the area of stereotyping, consider the relationship between accentuation bias and the failure to base typicality judgments on diagnostic ratios. As the correlation between trait attributions to men and women decreases (i.e., greater accentuation), the confound between trait attributions to the target group and the diagnostic ratio

increases. Thus, typicality ratings derived associatively from trait attributions become more similar to the typicality ratings that people would rationally derive from diagnostic ratios. The surprising result is that the less accurate participants' estimates are because of gender accentuation, the more coherent these estimates are. In our data, the size of the accentuation effect predicted even the unconfounded (i.e., unique) predictive power of the diagnostic ratio ( $M = .38$  and  $.22$  for the American and Italian sample, respectively; both  $ps < .01$ ).<sup>4</sup>

Social perceivers thus face a dilemma. When making trait attributions to a social group, they may strive for accuracy or coherence but will find it difficult to attain both. On one hand, associations between certain groups of people and their presumed psychological characteristics are easy to learn. They provide guideposts for action and they may be fairly accurate. On the other hand, these associations produce systematic contradictions, which can only be undone by controlled rational thought (Epstein, 1994). The dilemma is to know just when and how to step in to override automatic associations. At the current state of our understanding, we suspect that this question still falls into the ethical realm (Banaji & Bhaskar, 2000).

#### APPENDIX A Three Categories of Trait Terms

Masculine traits	
aggressive ( <i>aggressivo</i> )	assertive ( <i>sicuro di sé</i> )
defends beliefs ( <i>difende le opinioni</i> )	dominant ( <i>dominate</i> )
forceful ( <i>vigoroso</i> )	independent ( <i>independente</i> )
leadership ability ( <i>capacità di leader</i> )	strong personality ( <i>forte personalità</i> )
takes a stand ( <i>disposto a prendere posizione</i> )	takes risks ( <i>disposto a rischiare</i> )
Feminine traits	
affectionate ( <i>affettuoso</i> )	compassionate ( <i>compassionevole</i> )
gentle ( <i>delicato</i> )	loves children ( <i>amante dei bambini</i> )
sensitive ( <i>sensibile</i> )	soothing ( <i>consolante</i> )
sympathetic ( <i>empatico</i> )	tender ( <i>tenero</i> )
understanding ( <i>comprensivo</i> )	warm ( <i>caloroso</i> )
Neutral traits	
adaptable ( <i>adattabile</i> )	conscientious ( <i>coscienzioso</i> )
friendly ( <i>amichevole</i> )	happy ( <i>felice</i> )
helpful ( <i>disponibile ad aiutare</i> )	likeable ( <i>gradevole</i> )
reliable ( <i>fidato</i> )	tactful ( <i>che ha tatto</i> )
truthful ( <i>schietto</i> )	sincere ( <i>sincero</i> )

SOURCE: After Bem (1981).

NOTE: Italian translations are in italics.

**APPENDIX B**  
**Predicting the Natural Confound**  
**Between Attribution and Categorization**

The dependency of difference scores on trait attributions to the target group can be predicted from the correlation between the attributions to the two groups and their respective variances. The formula, modified after McNemar (1969, p. 177), is

$$r_{x, y-x} = \frac{s_x - r_{xy} s_y}{\sqrt{s_x^2 + s_y^2 - 2r_{xy} s_x s_y}}$$

For uncorrelated variables with the same variance, for example, the correlation between X and X-Y is .707. The dependency of the ratio X/Y on X is not derived as handily. To estimate it, we ran a simulation with 1,000 independent trials in S-PLUS (Becker, Chambers, & Wilks, 1996). Input values for X and Y ranged from 1 to 9 with each value drawn with the same probability. The simulation was validated in two ways. First, the distribution of the correlations between X and Y ( $M = .004$ ,  $SD = .185$ ) mapped onto the distribution expected from a chance model ( $M = 0$ ,  $SD = .192$ ). Second, the simulated correlation between X and X-Y ( $M = .712$ ) mapped onto the correlation obtained from McNemar's formula. Therefore, the simulated value for the correlation between X and X/Y ( $M = .550$ ) was a credible estimate.

Differences scores are thus more confounded with trait attributions to the target group than are ratios,  $t(999) = 78$ . The implication is that unless trait attributions to the target group are partialled, difference score measure will appear to provide the strongest support for the categorization hypothesis. Empirical research on national stereotypes has demonstrated this effect (Krueger, 1996b, Table 4).

In the simulation, the dependency increased as the correlation between X and Y became more negative, and this trend was stronger for difference scores ( $r = -.826$ ) than for ratios ( $r = -.615$ ). Also as expected, correlations between X and X-Y changed as correlations between X and X/Y did ( $r = .751$ ).

**NOTES**

1. The odds form of Bayes's Theorem shows the relationship between the diagnostic ratio and categorization. The categorization ratio is the probability that a person is a man given that Trait T is observed divided by the probability that the person is a woman given the trait. This ratio is identical to the diagnostic ratio times the base rate ratio, or

$$\frac{p(M|T)}{p(W|T)} = \frac{p(T|M)}{p(T|W)} \times \frac{p(M)}{p(W)}$$

Because the base rate ratio for the gender categories is close to 1, the diagnostic ratio expresses categorization directly. Because  $p(M|T) + p(W|T) = 1$ , the denominator of the categorization ratio carries no independent information. Therefore, the diagnostic ratio fully predicts the probability that a person with Trait T is a man.

2. The idiographic correlations between cultural stereotype ratings and personal typicality ratings remained positive when both the group averages of the cultural ratings and the group averages of the personal ratings were statistically controlled ( $M_{\text{American}} = .30$ ,  $M_{\text{Italian}} = .35$ ), replicating an analogous finding in the area of racial (African American vs. Caucasian American) stereotyping (Krueger, 1996a; see also Kenny & Winquist, 2001, for a discussion of the psychometric properties of this analysis).

3. When perceptions of two groups are at least somewhat accurate in that percentage estimates are positively correlated with the corresponding criterion percentages, correlations between estimated and actual diagnostic ratios regress toward zero. A simulation demonstrating this basic statistical property is available from the authors.

4. The measure of accentuation was each participant's Z-scored correlation between percentage estimates for men and percentage estimates for women computed across all 30 traits. The measure of predictive power was the average Z-scored partial correlation between the diagnostic ratio and typicality ratings with trait attributions to the target gender being controlled. The partial correlation was an average because each participant judged both genders.

**REFERENCES**

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

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology*, *74*, 955-966.

Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior*. Hillsdale, NJ: Lawrence Erlbaum.

Auster, C. J., & Ohm, S. C. (2000). Masculinity and femininity in contemporary American society: A reevaluation of the Bem Sex Role Inventory. *Sex Roles*, *43*, 499-528.

Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 139-175). Cambridge, MA: Harvard University Press.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 629-649.

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1996). *The new S language: A programming environment for data analysis and graphics*. New York: Chapman & Hill.

Bem, S. L. (1981). *The Bem Sex-Role Inventory: A professional manual*. Palo Alto, CA: Consulting Psychologists Press.

Beyer, S. (1999). The accuracy of academic gender stereotypes. *Sex Roles*, *40*, 787-813.

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

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

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluation and personality: Negative implications for mental health. *Journal of Personality and Social Psychology*, *68*, 1152-1162.

Conway, M., & Vartanian, L. R. (2000). A status account of gender stereotypes: Beyond communality and agency. *Sex Roles*, *43*, 181-199.

Dawes, R. M. (1970). An inequality concerning correlation of composites vs. composites of correlations. *Oregon Research Institute Technical Report*, *1*(1).

Dawes, R. W. (1998). Behavioral decision making and judgment. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 497-548). Boston: McGraw-Hill.

Diekmann, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*, 1171-1188.



- Diekmann, A. B., Eagly, A. H., & Kulesa, P. (2002). Accuracy and bias in stereotypes about the social and political attitudes of women and men. *Journal of Experimental Social Psychology, 38*, 268-282.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709-724.
- Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83*, 37-64.
- Ford, T. E., & Stangor, C. (1992). The role of diagnosticity in stereotype formation: Perceiving group means and variances. *Journal of Personality and Social Psychology, 63*, 356-367.
- Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 234-257.
- Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology, 77*, 350-359.
- Hasman, J. F., & Krueger, J. (2002). [The typicality of gender-descriptive attributes]. Unpublished raw data.
- Holt, C. L., & Ellis, J. B. (1998). Assessing the current validity of the Bem Sex-Role Inventory. *Sex Roles, 39*, 929-941.
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review, 100*, 109-128.
- Kenny, D. A., & Winquist, L. (2001). The measurement of interpersonal sensitivity: Consideration of design, components, and unit of analysis. In J. Hall & F. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 265-302). Englewood Cliffs, NJ: Lawrence Erlbaum.
- Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25*, 585-594.
- Krueger, J. (1992). On the overestimation of between-group differences. *European Review of Social Psychology, 3*, 31-56.
- Krueger, J. (1996a). Personal beliefs and cultural stereotypes about racial characteristics. *Journal of Personality and Social Psychology, 71*, 536-548.
- Krueger, J. (1996b). Probabilistic national stereotypes. *European Journal of Social Psychology, 26*, 961-980.
- Krueger, J. (2002). [Stereotype learning from frequencies]. Unpublished raw data.
- Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The contribution of social-perceptual skills and statistical regression to self-enhancement biases. *Journal of Personality and Social Psychology, 82*, 180-188.
- Krueger, J., Rothbart, M., & Sriram, N. (1989). Category learning and change: Differences in sensitivity to information that enhances or reduces intercategory distinctions. *Journal of Personality and Social Psychology, 56*, 866-875.
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*, 284-308.
- Lambert, A. J. (1995). Stereotypes and social judgment: The consequences of group variability. *Journal of Personality and Social Psychology, 68*, 388-403.
- Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology, 52*, 489-499.
- McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology, 36*, 929-940.
- McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York: John Wiley.
- Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality*. Oxford, UK: Blackwell.
- Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413-417.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.
- Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology, 32*, 309-336.
- Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70*, 1126-1141.
- Spears, R., Doosje, B., & Ellemers, N. (1997). Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification. *Personality and Social Psychology Bulletin, 23*, 538-553.
- Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology, 57*, 37-40.
- Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues, 25*, 79-97.
- Wilcox, C., & Francis, L. J. (1997). Beyond gender stereotyping: Examining the validity of the BEM Sex-Role Inventory among 16- to 19-year old females in England. *Personality and Individual Differences, 23*, 9-13.
- Williams, J. E., Satterwhite, R. C., & Best, D. L. (1999). Pancultural gender stereotypes revisited: The five factor model. *Sex Roles, 40*, 513-525.
- Zawadzki, B. (1948). Limitations of the scapegoat theory of prejudice. *Journal of Abnormal and Social Psychology, 43*, 127-141.
- Zuckerman, M., & Knee, C. R. (1996). The relation between overly positive self-evaluation and adjustment: A comment on Colvin, Block, and Funder (1995). *Journal of Personality and Social Psychology, 70*, 1250-1251.

Received October 10, 2001

Revision accepted May 29, 2002