

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Disentangling stereotype and person effects: Do social stereotypes bias observer judgment of personality?

Wayne Chan*, Gerald A. Mendelsohn

University of California, Department of Psychology, Institute of Personality and Social Research, 4143 Tolman Hall MC 5050, Berkeley, CA 94720-5050, United States

ARTICLE INFO

Article history: Available online 13 February 2010

Keywords: Person perception Individuation Stereotyping Observer judgment Impression formation

ABSTRACT

To what extent are observer judgments biased by social stereotypes? We calculated the relative contributions of stereotype and individuating information to personality judgment. Participants read several instant messaging conversations and rated the partners on trait scales. In the combined information condition, participants received individuating information from the conversation and stereotype information in the form of systematically varied gender-by-ethnic group labels. In the category information condition, participants rated "typical" members of each group, and in the transcript information condition, participants rated conversation partners without label. Person effects accounted for the majority of explained variance, and gender, but not ethnic, labels significantly affected the ratings. When abundant individuating information was available, the force of stereotypes, though not entirely eliminated, was significantly reduced.

© 2010 Elsevier Inc. All rights reserved.

JOURNAL OF RESEARCH IN PERSONALITY

1. Introduction

Apart from self-report inventories, there is no more important source of data in personality psychology than judgments made by human observers. Whether provided by a peer, a family member, an acquaintance or a research assistant these judgments can in Funder's words "be thought of as a kind of personality assessment, or even a personality test" (Funder, 2001, p. 122). It follows, he argues, that the psychometric considerations (e.g., reliability, response biases, and validity) that apply to other measuring devices apply equally to observer judgments. Funder also recognizes, however, that "as an instrument for gathering data, the human judge has some attributes that are distinctive ... distinctive attributes of the human judge include the way the distinction between the source of the data and the person who reads the data can become blurred and the way that many sources of bias possible with a human judge are different from those associated with more mechanical measures" (Funder, 1999, p. 7). It is inevitable, that is, that the wishes, goals, beliefs, intentions and expectations of the judges will affect and perhaps distort their judgments. By aggregation, the evaluation of inter-judge reliability and appropriate attention to the sampling of observers the effects of the blurring can be much reduced. But to the degree that biases, for example in the form of stereotypes, are widely shared in a population, the effectiveness of these steps will be limited. If the great majority of earthlings were in accord that Martians are cuddly and this belief had a strong influence on personality judgments made about individual Martians, aggregation and sampling would serve to delineate the stereotype, but would do nothing to improve accuracy. And, though inter-judge agreement would be high, there would be little differentiation among the targets with respect to stereotyped traits. In sum, we have means to deal with individual differences in observer biases, for example, by aggregation, but when biases are shared, the issues raised are of a different kind.

There is abundant evidence that attaching an ethnic or gender label to a target can have effects on a wide range of judgments. To give a simple example, identifying otherwise identical job applicants as female rather than male leads to lower ratings of competence (Correll, Benard, & Paik, 2007, Foschi, 1996) and to lower recommended salaries, as well (Smith, Tabak, Showail, Parks, & Kleist, 2005). To conclude that stereotypes affect judgments of others is not, however, to say that they determine those judgments. When there is no or minimal individuating information about the target, i.e., when group identification constitutes the only differentiating information available, it will necessarily be determining (Postmes, Spears, & Lea, 2002). Apart from studies of stereotypes and stereotyping and some anonymous Internet communications, such circumstances are rare. Typically judges have both kinds of information and must balance them in some way. One important approach to studying how the balancing takes place assumes that category-based (stereotype) and attribute-based (individuating) information are competing influences on person-perception (Neuberg & Fiske, 1987). Dual-process models propose that perceivers default to automatic use of stereotypes unless they recruit more controlled processes of perception (Devine, 1989), for example, when one is motivated to be non-prejudiced (Plant & Devine, 1998). To be sure, individual differences in the perception of a single target have been found such that people high in negative implicit associations to an outgroup label (Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003) or high in explicit negative attitudes towards an outgroup (Sherman, Stroessner, Conrey, & Azam, 2005) are more likely to perceive people in ways consistent with the label given. Typically, in experimental studies of this kind, participants receive a bit of behavioral information, mini vignettes or a set of adjective descriptors about a single target, and attentional processes are controlled or manipulated. Generalizing from these data to more everyday settings in which personality impressions are formed is difficult.

The objective of the study to be reported is to extend previous research by using a methodology that allows for the investigation of category and individuating information as conjoint rather than competing sources of influence on person perception. To make the experiment as naturalistic as possible and to enhance its contemporary relevance, we drew our stimulus materials from actual instant messaging conversations between pairs of students meeting for the first time and becoming acquainted with each other. Instant messaging permits participants to be anonymous while engaging in rich, complex, extended, real-time interactions. The text of the conversations was presented to observers who used multi-trait scales to give their personality impressions of the participants. Note that the judges saw just what the members of the dyads had seen as they were conversing. The conversations themselves contained no information about the gender or ethnicity of the participants; complete anonymity was maintained. In the experiment, group identity labels were systematically attached to the participants, as we will detail below. By providing the observers with both substantial individuating information and (experimentally manipulated) stereotype-relevant labels, the relative contributions of the two sources of information to personality impressions could be assessed. Although, as we noted above, the issue of how stereotypes affect personality judgments has been investigated in the past, studies comparable to the present one in which (1) there were multiple target persons and multiple labels. (2) abundant individuating information about the persons, (3) a reasonably naturalistic experimental setting using ecologically valid materials and (4) a comprehensive list of traits to be judged are not easy to find.

2. Overview

The basic strategy of the research was to obtain trait ratings of multiple persons each of whom was multiply labeled. Each of six persons from three dyadic conversations was assigned each of six ethnicity (African American, Asian American, and European American) by gender (male and female) labels in a six by six orthogonal design. There were thus 36 targets, each a unique combination of person and identifying label. This will be referred to as the "combined information" condition. By keeping person and label experimentally independent, the relative contributions of individuating information and stereotypes could be evaluated separately and interactively. In addition, two control conditions were run: in the category information control, judges described the "typical" member of each of the six ethnicity by gender groups, i.e., they were given the category label but had no individuating information. In the transcript information control, judges had individuating information, i.e., they read the conversations, but without any category labels attached. We used two complementary modes of approach to the data analysis. First, we examined the relative effect sizes of the person, the label and their interaction via a series of ANOVAs. Second we assessed the overall similarity of the trait ratings of the various targets using Q-correlations (Block, 1961/1978; Stephenson, 1953). Here, the key question was whether the coefficients of similarity for the same person variously labeled were less or greater than those of various persons assigned the same label.

3. Methods

3.1. Combined information sample

Psychology students (244) (age M = 20.20, SD = 2.44, 156 females, 82 males, six declined to state gender) at the University of California at Berkeley completed the experiment for partial course credit. The ethnic breakdown of the participants was as follows: six African American, 111 East Asian American, 62 European American, 14 Latino American, six Middle Eastern American, one Native American, 10 South Asian American, 22 Americans of mixed descent, and 12 declined to state. Participants were tested in groups of two to six. Upon arrival, they received a packet of three transcripts which they were told were conversations between two university students, each of whom was described as belonging to one of the six ethnicity × gender categories (e.g. "participant 3 is a European American male").

Each participant rated six targets. All six persons and all six labels were represented in each set of six. The targets were randomly assigned to participants, but with an essential constraint, namely that we presented each person and each label to a given participant once, and only once. A fully balanced design would have required each participant to rate all 36 targets, a logistically daunting option. Furthermore, once a person had been assigned a group label, it would have been unconvincing to assign a different group label to the same person for the same participant ("But I thought you said she was an African American male!"). However, by exposing each participant to all six persons and all six labels, we insured that any idiosyncratic rater tendencies should apply equally to each condition.

Participants were asked to read each transcript, and to form impressions of the targets. They completed ratings of each target's personality on a 31-item trait questionnaire. A 7-point Likert scale was used for the ratings. After completing the ratings for each of the six targets, participants filled out a short form reporting their own age, ethnicity, and gender.

To obtain participants' perceptions of the six persons free of stereotyping information, a subsample (N = 10) of participants completed the study as described above, except they were not supplied with any group membership information. This subsample will be referred to as the *transcript information sample*. Across the six persons, the mean inter-judge reliability was $\alpha = .78$ for this sample of 10.

3.2. Category information sample

In order to assess the relationship between the generated ratings and the prevalent stereotypes at the institution where the data were collected, 230 psychology students at the same institution in a separate sample (age M = 20.21, SD = 2.41, 127 females, 102 males, one declined to state gender) reported their impressions of "a typical member" of each of the ethnic \times gender groups using the same personality scales as above. The ethnic breakdown of the participants was as follows: six African American, 59 East Asian American, 100 European American, 13 Latino American, eight Middle Eastern American, 12 South Asian American, 29 Americans of mixed descent, and three declined to state. Data from the control samples were used to assess the similarity of the ratings of the targets to the ratings of the persons without group membership information, and to the stereotypes without individuating information.

4. Materials

4.1. Transcripts

Williams and Mendelsohn (2008) asked psychology students at the University of California at Berkeley to engage in dyadic "getting acquainted" conversations using a synchronous text-based medium (Yahoo Messenger). Participants in that study arrived singly in one of two labs (so that they never saw each other), and engaged in an anonymous online conversation for 15 min. The resulting conversations were recorded as text files, hereafter referred to as "transcripts." From the transcripts recorded in that study, the first author selected three of comparable length in which there was no explicit disclosure of gender or ethnicity. Those three transcripts, used verbatim and unedited, served as the stimuli in the current study. Since there were three transcripts each containing two participants, a total of six distinct persons (hereafter referred to as "persons") appeared in the transcripts.

4.2. Personality ratings

After reading the transcript of a conversation, participants rated the targets on a list of 31 trait terms, selected from the Adjective Check List (Gough & Heilbrun, 1983). The terms were chosen to provide a full range of social desirability, from very positive adjectives to very negative ones (Anderson, 1968; Gough & Heilbrun, 1983). This list was used by Williams and Mendelsohn (2008) in their study. The full list of adjectives follows in alphabetical order: affectionate, ambitious, argumentative, boastful, capable, coarse, cold, confident, dependent, egotistical, enterprising, feminine, flirtatious, good-natured, helpful, immature, inventive, irritable, logical, loyal, masculine, moody, self-centered, self-controlled, shy, submissive, sympathetic, unemotional, unselfish, warm, worrying. Each of these items was rated on a 7-point Likert-type scale, where one was "not at all", and seven was "very much." Williams and Mendelsohn (2008) grouped the trait terms into four scales: agreeable (affectionate, sympathetic, unemotional (reverse-scored), warm), agency (ambitious, capable, confident, enterprising), negative emotionality (irritable, moody, worrying), and assertive self-centered (argumentative, boastful, coarse, egotistical, immature, self-centered, unselfish (reverse-scored)). In our sample, the internal consistency reliabilities of the four scales were .74, .71, .65, and .82, respectively. We used these scales in our analyses, and also analyzed separately the three trait terms that produced the largest group differences in the stereotype data: masculine, feminine, and shy. Thus, there were seven dependent variables in the study: the four scales and the three single adjectives.

5. Results

Two aspects of the design create potential problems of nonindependence. First, each participant gave six ratings, and second, targets were always observed in interaction with a partner, who was also a target. We addressed the first problem by controlling for individual differences in scale usage by participants. We obtained the mean and standard deviation of the six ratings made by each participant for each dependent variable, and converted each observation to a *Z*-score, i.e., we ipsatized each participant's scores on each dependent variable (e.g., Cattell & Brennan, 1994). The same procedure was followed in the category, transcript, and combined information conditions. The second problem arises from the dyadic nature of our data, both because there were two persons in conversation, and because one rater makes judgments of both persons. Following the recommendations of Kenny (1988), we first examined the intraclass correlations (Shrout & Fleiss, 1979) between

W. Chan, G.A. Mendelsohn/Journal of Research in Personality 44 (2010) 251-257

the two targets in each dyad. Those analyses revealed that the personality ratings of the two targets in each dyad were essentially uncorrelated (Shrout-Fleiss reliability, random set, all ICC(2, 1) < .10). Given the empirical independence of the ratings within pairs, data were subsequently analyzed at the person level. In summary, the analyses reported for the experimental sample were 3-way, between cells, 6 (persons) \times 2 (gender) \times 3 (ethnicity) ANOVAs, with between 36 and 43 participants per cell.

5.1. Category information condition findings

There were significant gender and ethnic effects for all of the dependent variables. On the left side of Table 1, we list the effect sizes, F values and significance levels for the gender and ethnic stereotype effects by variable. It is clear that, even though the data were collected in an alleged bastion of political correctness, ethnic and gender stereotypes are alive and well in our sample.

5.2. Combined information condition findings

When, as in this study, ample individuating information is available, will gender and ethnic information continue to affect personality judgments? Or will individuating information overcome or attenuate the effect of labels? Analytically, this question boils down to assessing the relative influence of group labels and persons on the dependent variables. Toward this end, each of the seven ipsatized dependent variables was subjected to the $6 \times 3 \times 2$ between-participants ANOVA described above.

5.3. Effects of group labels

Gender labels significantly affected ratings of agreeable, masculine, and feminine. Ethnic labels significantly affected ratings of agreeable, assertive self-centered, agency, masculine, feminine, and shy. On the right side of Table 1, we list the effect sizes, *F* values and significance levels for the gender and ethnic label effects by variable. No significant ethnic *X* gender interactions were found. It is important to note that for the variables yielding a large effect ($\eta^2 > .14$; Cohen, 1988) in the analyses of the category information data, the same ordering of the means was obtained in the combined information conditions.¹ Thus, even when copious individuating information is available, a significant influence of stereotypes on personality judgments can still be observed.

5.4. Effects of persons

Participants significantly differentiated the six persons on all dependent variables. The middle column of Table 1 lists the person effect by variable. None of the two-way or three-way person \times group level interaction effects was significant. That is, raters applied the category information consistently across the persons, and described the six persons consistently however they were labeled.

Do these effects hold regardless of the gender and ethnicity of the rater? Does rater group membership and target group membership interact in any meaningful way? To address this question of group differences, we added two variables (rater gender, and rater ethnicity) to the previous three, making it a $6 \times 3 \times 2 \times 2 \times 2$ 5-way between cells analysis of variance. Due to the low number of participants of other ethnicities, we used only Asian American and European American participants in these analyses. Of focal interest were the gender label *X* participant gender interaction, the ethnic label *X* participant ethnicity interaction,

¹ The pattern of means was for agreeable: female > male, masculine: M > F, feminine: F > M, assertive self-centered: European American > Asian American > African American, masculine: AfAm > EuAm > AsAm, shy: AsAm > EuAm > AfAm.

Table 1

Effect sizes (η 2) and F values of the label and person effects in the category information and combined information conditions.

	Category information condition		Combined information condition			
	Gender label effect	Ethnic label effect	Person effect	Gender label effect	Ethnic label effect	
Agreeable						
F	260.01***	16.90****	43.34***	8.68**	3.91*	
η^2	.16	.02	.13	.01	.00	
Assertive self-centered						
F	151.37***	203.68****	33.70***	.98	10.86***	
η^2	.08	.21	.10	.00	.01	
Agency						
F	85.68***	5.58**	43.92***	.43	4.83**	
η^2	.06	.01	.13	.00	.01	
Negative emotionality						
F	62.56***	35.33***	7.74***	2.37^{+}	3.82*	
η^2	.04	.05	.03	.00	.00	
Masculine						
F	1383.94***	296.13***	33.82***	616.81***	16.68***	
η^2	.42	.18	.08	.27	.01	
Feminine						
F	1381.05***	121.05****	25.55***	548.06***	8.76***	
η^2	.46	.08	.06	.26	.01	
Shy						
F	14.06***	534.38***	5.48***	.87	6.44*	
η^2	.01	.44	.06	.00	.01	

⁺ *p* < .10.

p <.001.

and the 4-way gender label *X* ethnic label *X* participant gender *X* participant ethnicity interaction. Analyses of all of the dependent variables revealed no significant effects for any of the three focal interactions that would indicate in-group bias; there was no evidence of systematic variations of judgments as a function of group membership.

5.5. Relative contribution of group labels and targets

We address next the question of partitioning the explained variance in the combined information condition into its person and label components. The proportions of the between cells sums of squares attributable to the person, the gender label, the ethnic label and the interaction effects were calculated. As shown in Fig. 1, for most variables, the majority of explained variance was



Fig. 1. The relative contributions of person, gender label, ethnic label, and the interaction terms to explained variance in the combined information sample.

attributable to person. Masculine and feminine were exceptions, though person still accounted for a sizable portion of the explained variance of each. Not unexpectably, the majority of the explained variance for masculine and feminine was associated with gender labels. In no case did ethnic labels account for more than a small fraction of the variance accounted for by person effects. Finally, to compare the strength of labeling effects in the category information and the combined information condition we tested the difference between the two effect size estimates as recommended by Rosenthal and Rosnow (2008). In every case, the effect size of the label was reduced in the presence of individuating information. Inspection of Table 2 reveals that when individuating information was present, the effect size of labels was reduced significantly. Indeed, in the only two cases where individuating information did not result in a significant reduction of the label effect (the ethnic effect on agency, and the gender effect on shy), the association between the trait and the category was very weak ($\eta^2 = .01$).

5.6. Similarity ratings

The analyses just reported were variable-centered in the sense that the partitioning of the variance into its person and label

Table 2

Effect size reductions (Z of difference) between the category information and combined information conditions for gender and ethnic effects.

	Gender label	Ethnic label
Agreeable	9.20***	2.00*
Assertive self-centered Agency	7.00 6.14 ^{***}	.31
Negative emotionality	4.15***	4.36***
Masculine	4.94***	8.66***
Feminine	7.13***	5.33***
Shy	1.38	18.59***

* p < .05.

**** p < .001.

^{*} *p* < .05.

p < .01.

components proceeded variable by variable through the set of seven dependent variables. In the next series of analyses, we followed a person (target)-centered approach that takes each target's full profile of trait ratings as the unit of analysis. By correlating the arrays of ratings, a metric of similarity (Q-correlations) could be obtained. To be more specific: (1) since each target was rated by several judges, the means of all ratings for each target on each trait were obtained; (2) we calculated the Q-correlations between the targets, i.e., we correlated the rows, rather than the columns, of the standard subjects by variables matrix, where the subjects were the 36 rows and the columns were the trait terms. Thus we obtained, for example, the product-moment correlation of person 1 labeled as an African American male with person 1 labeled as an Asian American male, and so forth. Of central interest was the question of whether personality profiles of the same person given different labels were more similar to each other (i.e., had higher O-correlations) than personality profiles of different persons given the same group label. It is important to note that in calculating the Q-correlations, we did not include both masculine and feminine, which are negatively correlated in excess of .70. To eliminate this redundancy but retain the trait dimension, only feminine was retained in the Q analyses.

Because Q-correlations do not have a known distribution, we constructed a distribution of correlation coefficients using all possible pairs of targets. (See Klohnen and Mendelsohn (1998) for a similar approach.) In all, there were 630 correlations. We applied Fisher's *r*-to-*z* transformation to each of the correlations, and in all the following Q analyses, descriptive statistics and inferential tests were based on the z-transformed correlations. The reported rs and SDs below were back-transformed from the Fisher's zs. The mean Q-correlation for all pairs was r = -.02 (SD = .42, range -.83 < r < .87). We then compared the distribution of Q-correlations for a class of pairs (e.g. the same person with different labels) to the distribution of the remaining Q-correlations. The mean Qcorrelation between personality profiles of the same person was r = .55 (SD = .35), significantly higher than the mean Q-correlation for pairs that do not contain the same person (r = -.12, SD = .35, t(628) = 17.62, p < .001, d = 1.91; in sum, participants rated a person given one label as more similar to the same person given a different label than to a different person however labeled. In contrast, the mean Q-correlation of two different persons given the same ethnic label (r = -.01, SD = .39) was not significantly different from the mean Q-correlation of the remaining pairs (r = -.02, SD = .43, t(628) = .16, ns, d = .02). Correlations of personality profiles of two persons given the same gender label (r = .11, SD = .41), however, were significantly higher than those of the remaining pairs, (r = -.14, SD = .40, t(628) = 7.19, p < .001, d = .62).² It is clear, then, that while individuating information in the transcript had more of a role in coloring perception than the group label information, group label information, specifically gender labels, also significantly altered person perception. The results at the level of the profile are consistent with the results at the level of the personality variables in showing the centrality of the person in the formation of personality impressions, with gender stereotyping playing a significant secondary role.

5.7. Similarity of target to control condition descriptions

The last analyses to be reported involved comparisons of the trait profiles obtained in the combined information condition to the profiles obtained in the category information and transcript information conditions. After calculating the item means for each of the six persons in the transcript information condition, we obtained the Q-correlations between each person's control profile of means and the profile of means of the 36 targets. Of the resulting 216 correlations, 36 were within person, i.e., the person in the transcript information condition was the same as the (labeled) person in the combined information condition. The mean correlation within person was .48 (SD = .33) and that across persons was -.09 (SD = .35), t(214) = 9.87, p < .001, d = 1.68. There was significant similarity between the control descriptions of the six persons without label and the descriptions of those persons (variously labeled) in the combined information condition. In parallel fashion, we assessed the similarity of the personality ratings of the targets to the stereotypes obtained in the category information condition. We conducted a 2×2 (same vs. different gender \times same vs. different-ethnicity) ANOVA, where the Q-correlations were the dependent variable. Significant results were obtained both for gender (same-gender mean r = .18, SD = .41; different-gender mean r = -.19, SD = .41; F(1, 212) = 40.38, p < .001, d = .90) and for ethnicity (same-ethnicity mean r = .11, SD = .45; different-ethnicity mean r = -.06, SD = .42; F(1, 212) = 7.71, p < .01, d = .39). Both of these positive same label correlations demonstrate that participants rated the targets in the direction of the stereotypes. There was no significant gender X ethnicity interaction (F(1, 212) < 1,ns). Thus, these analyses make it clear that even when a large amount of individuating information is available and that information plays a predominant role in the perception of the targets, impression formation is nevertheless subject to the biasing effects of stereotypes.

6. Discussion

The distinctive feature of this research is the systematic analysis of the relative contributions of stereotype and individuating information to impression formation. The existence of stereotypes in our sample is clear: in the absence of any information other than a group label, i.e., when asked to describe a "typical" member of gender and ethnic groups, participants were in significant agreement about the distinguishing characteristics of the groups. Now, what happens when another source of information, a source that provides considerable information about the persons observed, is also available? How will the two sources of information combine and interact? Will the impressions formed of the targets still be affected by the labels attached to them? Will the influence of the stereotypes be nullified or diminished when they are attached not to the "typical" member of a group but to a particular person from that group? Answers to those questions are not simple, for, as shown in the presentation of results, they depend in part on the class of label (gender or ethnicity) and in part on the dependent variable of concern.

The data make it clear that, in general, the individuating information exercised a strong, directing influence on the descriptions of the targets. Significant (p < .001) effects of person were obtained for all the dependent variables and those effects were not moderated by label, i.e., the six persons even when labeled differently were described consistently. Examination of Fig. 1 shows that the majority, as high as 90%, of the explained variance of the four trait composites and of shy was attributable to person. The percentages were substantially lower for masculine and feminine, but still were at about 20%.

Despite the strength and extent of the person effects, there were also significant label effects for all but one of the dependent variables in the combined information condition. Examination of Table 1 shows that the it was those variables with the largest effect

² These data were based on a person × gender label × ethnic label $2 \times 2 \times 2$ ANOVA (with each variable dummy coded 1 for same, and 0 for different) which revealed significant person effects, *F*(1, 622) = 145.88, *p* < .001, and gender effects, *F*(1, 622) = 59.70, *p* < .001. There was a marginal ethnic effect, *F*(1, 622) = 3.04, *p* < .10. Importantly, there were no significant interactions among any of the variables.

sizes in the category information condition that produced the strongest gender and ethnic differences in the combined information condition, i.e., across the dependent variables, there was a direct relationship between the effect sizes in the category information and the combined information conditions. This is most clearly exemplified by the results for masculine and feminine, the two variables that were by far most strongly associated with gender in the stereotype data ($\eta^2 > .40$). In each case, large effect sizes were likewise found in the combined information condition and gender accounted for the great majority, about 70%, of the explained variance. In contrast, for those variables where gender and ethnicity had only a moderate or small effect size ($\eta^2 < .06$) in the stereotype data, η^2 in the experimental condition did not exceed .01 and labels accounted for no more than 10% of the explained variance. In sum, for any group label, there may be strong associations with some particular traits and weaker associations with others. Hence in a situation where judges assess targets, the likelihood that gender or ethnic biases will affect ratings of persons is a function of the strength of the association between the label and the given trait. This is consistent with existing research on judgments of behaviors of single targets (e.g. Krueger & Rothbart, 1988), in that categories more diagnostic of a trait have stronger effects. When, as with masculine and feminine, the association (both in self and in other ratings) to gender is so strong, person yields to label in accounting for the variance of ratings. When the association is less strong, person dominates decisively, significant label effects notwithstanding. These findings support our contention that stereotypes and individuation are not necessarily competing processes. In a reasonably naturalistic setting with sufficient personalizing information available, both contribute - in varying degrees-to the formation of impressions of personality.

Using Q-correlations as an additional approach to the analysis of the data had two distinct advantages. First, it utilized the full set of traits rather than proceeding from variable to variable. We were thus able to obtain similarity matrices between trait profiles that permitted an evaluation of the contributions of person and label to the overall descriptions of the targets. Second, we could more directly utilize the data of the category information and transcript information conditions by obtaining the correlations between the control profiles and the profiles of the targets in the combined information condition. These similarity coefficients were then used to address the questions of (1) whether a given person though variously labeled remained similar to the relevant control description and (2) whether different persons with the same label were seen as equally similar to the stereotype for that label. By its very nature, however, Q analysis is uninformative about particular variables. Thus the two analytic approaches are complementary not redundant. From the ANOVAs, we learned that the proportions of explained variance attributable to person were greatest for the trait composites and shy, whereas it was gender that accounted for the majority of the explained variance of masculine and feminine. These variable to variable differences are informative about the relationship of stereotype to biases in trait judgments, but any attempt to characterize the findings as a whole can be done only discursively. The Q analyses, on the other hand, do permit us to make a summary statement, namely that for the data set as a whole, all three sources of information color perception, but person matters most, gender makes an important secondary contribution and ethnicity matters to a small but still statistically significant degree.

The direct influence of individuating information on impression formation is evident in both sets of analyses. There is, however, another way, besides variance accounted for or relative magnitude of *Q*-correlations, in which the person matters. When labels were attached to particular persons rather than to the "typical" member of a group, the force of the label was sharply reduced. As shown in Table 2, for all but two of the dependent variables the label effect size was significantly smaller in the combined information than in the category information condition. This was the case even for masculine and feminine, the variables showing the strongest label effects. As for the other variables, none of the effect sizes in the combined information condition exceeded .02. It appears that behavior, in effect, engulfs the field and participants anchor on the individual in making their judgments. The individuating information, that is, serves to attenuate the stereotype. Still, we must recognize that the stereotypes were not entirely neutralized.

Before discussing the implications of these findings for the questions about observer reports that motivated the research, we note three limitations of the experimental design, all of which suggest profitable ways to extend the present study. First, the data were obtained from passive observers; they read transcripts of actual conversations but were not themselves active participants in a social interaction with the targets. True, they saw the same text as active participants did, in fact, see, but the psychological and interactional positions of active and passive observers differ. Most importantly, since active participants can to some extent control and direct the conversation, the presence of identifying labels could affect the content and flow of the exchange. The use of a passive observer design made it possible to establish the experimental control needed to separate person and stereotype effects, but further research exploring the influence of labeling on actual, realtime conversations is clearly indicated as a logical next step. On the other hand, the present research generalizes to the growing number of interactions on the Internet, in which one makes social judgments on the basis of observing an exchange between two people with no additional background other than basic category membership information.

The second limitation stems from the all or none fashion in which the individuating information was presented to participants; either they read the entire transcript or they did not. Consequently, we cannot assess at which point in an interaction individuation becomes the predominant factor in person perception. Based on our results, we would expect the answer to differ by trait. Whereas it may take only a thin slice of a person's behavior to make a judgment of the target's agreeableness, it may take considerably more exposure for a judge to conclude that the target is, say, an unusually masculine female (cf. Funder & Colvin, 1991; Funder & Dobroth, 1987; John & Robins, 1993). The systematic examination of the question of how much and what kind of individuating information is necessary to alter the balance between category and attribute-based processes in impression formation would be a useful focus of future research.

Thirdly, we did not assess possible effects of individual differences in our participants on the ratings of the targets. Are there variables that can predict who would be more likely to make use of individuating information in person perception? Such individual differences as openness to experience (Flynn, 2005), implicit person theories (Levy & Dweck, 1998), and susceptibility to stereotypes (Chan & Mendoza-Denton, 2008; Mendoza-Denton, Downey, Purdie, Davis, & Pietrzak, 2002) may affect participants' use of, and reactions to, social stereotypes. Future research should address the role that individual differences may play in the differential use of various sources of information in personality judgment.

No approach to the measurement of personality is immune from methodological problems of one kind or another. Our results indicate that Funder's (1999) concerns that the human judge, by virtue of being human, is susceptible to biasing factors not present in other kinds of assessment instruments is justified. Stereotypes, at least gender stereotypes, have an effect on descriptions of personality even when individuating information is abundant. Overall, however, the pattern of findings should be reassuring to the diverse group of researchers who regularly make use of observer ratings. There was no evidence that ethnic stereotypes had more than a minor influence on impression formation, none of the interactions between label and person were significant and there was not even a hint of an in-group bias in the trait ratings. We found instead that persons were well and consistently differentiated irrespective of how they were labeled. In sum, human judges given the chance to do so focused on the behavioral data and responded accordingly.

Acknowledgments

Wayne Chan and Gerald A. Mendelsohn, University of California, Berkeley. This research was supported by a National Science Foundation Graduate Research Fellowship to the first author. We gratefully acknowledge Kiik Araki-Kawaguchi, Stacey N. Cho, Melissa Hackel, Adrienne Man, Seema Sood, Diane Velasco, Joanne Wong, and Dan Yun for their assistance with data collection and data entry. In addition, we thank Elizabeth Page-Gould, Melissa J. Williams, Thomas D. Wickens, and anonymous reviewers for helpful comments on earlier revisions of this article. Electronic correspondence should be addressed to Wayne Chan at wychan@ berkeley.edu.

References

- Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. Journal of Personality and Social Psychology, 9, 272–279.
- Block, J. (1961/1978). The Q-sort method in personality assessment and psychiatric research. Palo Alto, CA: Consulting Psychologists Press.
- Cattell, R. B., & Brennan, J. (1994). Finding personality structure when ipsative measurements are the unavoidable basis of the variables. *American Journal of Psychology*, 107, 261–274.
- Chan, W., & Mendoza-Denton, R. (2008). Status-based rejection sensitivity among Asian Americans: Implications for psychological distress. *Journal of Personality*, 76, 1317–1346.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? American Journal of Sociology, 112, 1297–1338.
- Devine, P. G. (1989). Stereotypes and prejudice. Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Flynn, F. J. (2005). Having an open mind: The impact of openness to experience on interracial attitudes and impression formation. *Journal of Personality and Social Psychology*, 88, 816–826.
- Foschi, M. (1996). Double standards in the evaluation of men and women. Social Psychology Quarterly, 59, 237–254.

- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Funder, D. C. (2001). The personality puzzle. New York, NY: W.W. Norton.
- Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, 60, 773–794.
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 54, 409–418.
- Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology*, 39, 16–30.
- Gough, H. G., & Heilbrun, A. B. (1983). The adjective check list manual. Palo Alto, CA: Consulting Psychologists Press.
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61, 521–551.
- Kenny, D. A. (1988). The analysis of data from two-person relationships. In S. W. Duck (Ed.), Handbook of personal relationships. London: John Wiley & Sons.
- Klohnen, E. C., & Mendelsohn, G. A. (1998). Partner selection for personality characteristics: A couple-centered approach. Personality and Social Psychology Bulletin, 24, 268–278.
- Krueger, J., & Rothbart, M. (1988). Use of categorical and individuating information in making inferences about personality. *Journal of Personality and Social Psychology*, 55, 187–195.
- Levy, S. R., & Dweck, C. S. (1998). Trait- versus process-focused social judgment. Social Cognition, 16, 151–172.
- Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A., & Pietrzak, J. (2002). Sensitivity to status-based rejection: Implications for African American students' college experience. *Journal of Personality and Social Psychology*, 83, 896–918.
- Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, 53, 431–444.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. Journal of Personality and Social Psychology, 75, 811–832.
- Postmes, T., Spears, R., & Lea, M. (2002). Intergroup differentiation in computermediated communication: Effects of depersonalization. *Group Dynamics: Theory, Research, and Practice, 6*, 3–16.
- Rosenthal, R., & Rosnow, R. L. (2008). Essentials of behavioral research: Methods and data analysis. New York, NY: McGraw-Hill.
- Sherman, J. W., Stroessner, S. J., Conrey, F. R., & Azam, O. A. (2005). Prejudice and stereotype maintenance processes: Attention, attribution, and individuation. *Journal of Personality and Social Psychology*, 89, 607–622.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 2, 420–428.
- Smith, F. L., Tabak, F., Showail, S., Parks, J. M., & Kleist, J. S. (2005). The name game: Employability evaluations of prototypical applicants with stereotypical feminine and masculine first names. Sex Roles, 52, 63–82.
- Stephenson, W. (1953). The study of behavior: Q-technique and its methodology. Chicago, IL: University of Chicago Press.
- Williams, M. J., & Mendelsohn, G. A. (2008). Gender clues and cues: Online interactions as windows into lay theories about men and women. Basic and Applied Social Psychology, 30, 278–294.