Sex Differences in Judgments of Physical Attractiveness: A Social Relations Analysis
David K. Marcus and Rowland S. Miller
DOI: 10.1177/0146167202250193

The online version of this article can be found at:
http://psp.sagepub.com/content/29/3/325
Sex Differences in Judgments of Physical Attractiveness: A Social Relations Analysis

David K. Marcus
Rowland S. Miller
Sam Houston State University

Men and women rated the physical attractiveness of other men and women who were sitting nearby and were rated by them in return. They also provided meta-perceptions of how they thought those others rated them. Attractiveness ratings were partly a function of both the target being rated and the perceiver providing the ratings regardless of the sex of the perceiver or target, but the highest levels of consensus occurred when men judged the attractiveness of women and the highest levels of idiosyncrasy occurred when men rated other men. Meta-perceptions were also idiosyncratic; some believed that they were consistently considered attractive, whereas others thought they were seen as unattractive. People were aware of what others thought of them and, in particular, women’s meta-perceptions were highly related to men’s judgments of them. People agree about others’ attractiveness, and those who are attractive to others know they are pretty or handsome.

Keywords: physical attractiveness; sex differences; interpersonal perception

For better or for worse, a person’s looks matter. People who are judged to be physically attractive make generally more favorable impressions on others than do people with lesser looks. Strangers assume that attractive people have agreeable personalities, are more sociable, and are destined to be more desirable parents and employees than are those who are unattractive (Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992). Their criminal actions engender less official concern (Downs & Lyons, 1991). And, of course, they are coveted as romantic partners: both men and women typically desire as mates the most attractive partners they are able to win (Walster, Aronson, Abrahams, & Rottman, 1966).

Moreover, people generally agree about who is attractive: Consensus among judges is a robust finding (Jackson, 1992). A recent meta-analysis of studies in which perceivers rated the attractiveness of adults yielded high effective reliabilities regardless of whether the studies examined within-culture agreement ($r = .90$), cross-ethnic agreement ($r = .88$), or even cross-cultural agreement ($r = .83$) (Langlois et al., 2000). Obviously, beauty is not just “in the eye of the beholder;” such agreement among disparate judges suggests that some people seem physically attractive to almost anyone.

Indeed, certain physical characteristics are universally considered to be attractive. Appealing faces are symmetrical (Mealey, Bridgstock, & Townsend, 1999; Shackelford & Larsen, 1997) and combine neonate features such as large eyes and a small nose (Jones, 1995; Perrett et al., 1998) with relatively mature characteristics such as thin cheeks and high cheekbones (Cunningham, Roberts, Barbee, Druen, & Wu, 1995). Attractive faces also tend to have features that represent the mathematical average of the faces in a given population (Langlois, Roggman, & Musselman, 1994; Rhodes, Sumich, & Byatt, 1999). In addition, some bodies are more attractive than others. For instance, a narrow waist-to-hip ratio that presents an “hourglass” figure is more attractive in women than is a stouter figure in which the circumference of the waist is more similar to that of the hips (Singh, 1993; Singh & Luis, 1995). These features and characteristics (e.g., neonate vs. average features) may be linked to competing theories of attractiveness by their proponents, but the literature is clear: Some characteristics are universally attractive.

Authors’ Note: We thank Nazneen Askari, Patricia Vargas, and Elizabeth Link for their assistance with data collection. Correspondence concerning this article should be addressed to David K. Marcus, Department of Psychology, Sam Houston State University, Huntsville, TX 77341-2447; e-mail: psy_dkm@shsu.edu.

Prior studies have examined how consensus in judgments of attractiveness varies (or does not vary) with the ethnicity and nationality of the perceiver and target, but considerably less attention has been devoted to examining whether consensus varies with participants’ sex. Men care more about a partner’s looks than do women (Buss, 1999; Feingold, 1990), but Langlois et al. (2000) found that a target’s sex did not moderate the reliability of attractiveness ratings. However, these null findings may be due to limitations in the research reviewed by Langlois et al.: Most studies examined ratings of only male or female targets, and few studies compared the ratings provided by male and female perceivers when both sexes did participate. Unlike prior studies, the primary aim of the present study was to explore the ways in which sex may moderate the degree of interjudge agreement in perceivers’ ratings of the attractiveness of others. In other words, is there greater consensus when female targets are rated or when male targets are rated? Do male and female perceivers yield differing levels of consensus? Most important, do the sex of a target and the sex of a perceiver interact so that, for example, there are higher levels of consensus when perceivers rate members of the other sex?

Prior studies have not directly examined whether perceiver and target sex moderate consensus, but Graziano, Jensen-Campbell, Shebilske, and Lundgren (1993) did examine men’s and women’s perceptions of how expert they are at evaluating the physical attractiveness of other men and women. According to these students’ ratings (which probably reflected commonsense beliefs about gender differences), men and women agreed that women are more expert at judging the attractiveness of men. However, whereas men believed that they were better than women at judging the attractiveness of female targets, women believed that they were just as skilled as men at judging female attractiveness. Essentially, the men in this study predicted an interaction between the sex of the perceiver and the sex of the target in judgments of attractiveness, but women predicted a main effect for perceiver sex, believing themselves to be more expert in evaluations of others’ looks.

The question of whether sex moderates the degree of consensus in attractiveness ratings has important theoretical consequences because, as Langlois et al. (2000) suggested, different theories yield different hypotheses about the influence of sex on perceptions of attractiveness. Because our culture values female attractiveness more than male attractiveness (Owen & Laurel-Seller, 2000; Sobal, 1995), sociocultural theories (e.g., Eagly & Wood, 1999; Jackson, 1992) would presumably predict that the sex of the target should moderate consensus, with greater consensus about the attractiveness of women than men. With cultural influences socializing both men and women to pay particular attention to women’s looks, a sociocultural perspective offers little reason to expect the sex of the perceiver to influence the degree of consensus.

In contrast, an alternative perspective that emphasizes the mechanism of mate selection in the evolution of attractiveness preferences (e.g., Berry, 2000; Buss, 1999; Cunningham, 1986) predicts that men and women should differ in the precision with which they judge women’s looks. According to this view, physical attractiveness is a crude but useful marker of reproductive fitness in women (Singh, 1993), and men who astutely judged the looks of potential partners may have reproduced more successfully than did those who were less attentive or discerning. In particular, because women’s fertility (unlike men’s) ends in middle age, it would have been more adaptive for men (than for women) to be vigilant for signs of youthful beauty in potential mates (Kenrick & Keefe, 1992). Attractiveness is an indicator of health and heritable fitness in men as well as in women (Singh, 1995), so an evolutionary perspective (e.g., Gangestad & Thornhill, 1997; Shackelford & Larsen, 1999) predicts an interaction of sex of perceiver and sex of target in which both sexes exhibit higher levels of consensus when rating the other sex than when rating their own. Nevertheless, this perspective also expects even greater agreement among men than among women in other-sex judgments and, unlike a sociocultural perspective, it predicts that men will agree in their judgments of women to a much greater degree than do women. Furthermore, consistent with an evolutionary perspective, there is evidence that women’s judgments of male attractiveness vary depending on the phase of their menstrual cycles (Penton-Voak & Perrett, 2000; Penton-Voak et al., 1999), also suggesting that there will be less consensus among female perceivers (who are presumably at varying points in their cycles) when rating men than among male perceivers when rating women.

Admittedly, the ultimate question from an evolutionary perspective is preference and not perception: Women may display as much consensus in their judgments of men as men do in their judgments of women, but women’s perceptions of men may not play as great of a role in their mate selection as men’s perceptions of women. Thus, the absence of sex differences in judgments of attractiveness may not be fatal to an evolutionary approach, but evidence of sex differences provides support for such a perspective.

Most studies that have examined the question of consensus have used procedures in which participants inspect photographs or, occasionally, videotapes of those whose attractiveness they are asked to judge (e.g.,
Bazzini & Shaffer, 1999; Lydon, Meana, Sepinwall, Richards, & Mayman, 1999). In fact, in their meta-analysis, Langlois et al. (2000) had to collapse video and in situ studies into a single category because there were so few in situ studies on record. However, in their daily transactions, in the judgments that matter, people appraise the looks of others who are standing in front of them; they have dynamic and complex facial and bodily cues available to them, and they evaluate the attractiveness of both men and women who, importantly, may be evaluating them in return. Furthermore, some of the most influential studies of attractiveness (e.g., Walster et al., 1966) have asked participants to rate the attractiveness of live individuals who were actually present. Following this tradition, we attempted to capture the real workings with which naturalistic judgments of others’ beauty are made.

This may be no small matter because studies with live targets may be more likely than those with photographic stimuli to yield sex effects like those we examine here. The difference between men’s motivation to judge carefully the attractiveness of women and their motivation to judge other men (or women’s motivation to judge men compared to their motivation to judge other women) may be greatest when the targets are present and potentially available as romantic partners. Any interaction between perceiver sex and target sex is likely to be smaller when the targets are photographs of people whom the raters will never actually meet. Because so few existing studies have employed (a) both men and women as (b) both perceivers and targets of (c) judgments of live peer models, we know of no prior study that investigated the issues we examine here.

Moreover, in the field, perceptions of attractiveness ordinarily involve comparative judgments of multiple targets of both sexes who are often simultaneously evaluating each other. One goal of attractiveness research should be to model these complex, interactive judgments, which often take the form of natural “round-robin” designs in which people assess the looks of multiple targets while they are being judged by each of the targets in return. Happily, Kenny’s (1994) Social Relations Model (SRM) is specifically designed to make sense of such situations.

The Social Relations Model

Because the SRM is becoming better known (e.g., Kenny, 1994; Malloy, Sugarman, Montvilo, & Ben-Zeev, 1995), our description of the model will be brief; interested readers may consult Kenny (1994) for a more thorough description of the model. The three principal components of the SRM are the target of an interpersonal judgment, the perceiver, and their relationship. As an example, suppose that Bill finds Monica very attractive. If Bill’s perception is widely shared and people generally find Monica attractive (i.e., there is consensus among diverse perceivers), there would be a large “target” effect: Monica is evidently quite good-looking. On the other hand, if Bill is not very discriminating and rates everyone as attractive, there would be a large “perceiver” effect. Finally, if Bill does not find most people attractive but thinks Monica is gorgeous when others generally do not, there would be a large “relationship” effect, indicating that an interaction of perceiver and target is influencing the result. When only one judgment of Monica by Bill is obtained, the SRM cannot distinguish a relationship effect from error variance; multiple ratings are necessary to separate the two. Still, when multiple judges rate multiple targets, the SRM determines the proportions of the variance in judgments of attractiveness that can be attributed to the targets, the perceivers, and relationships/error. Note that large amounts of target variance indicate that there is consensus among the judges. Substantial perceiver variance is indicative of assimilation (Kenny, 1994): From the perceiver’s point of view, diverse targets generally look the same.

Two prior social relations model studies (Albright, Kenny, & Malloy, 1988; Kenny, Horner, Kashy, & Chu, 1992) that examined consensus for judgments of personality traits among strangers also asked their participants to rate one another’s attractiveness. All three variance components significantly contributed to these attractiveness ratings, with perceivers accounting for 33% of the variance (indicating that assimilation occurred and perceivers had idiosyncratic preferences), targets accounting for 27% of the variance (consensus existed and diverse perceivers agreed about the attractiveness of particular targets), and their relationships accounting for 40% of the variance (Albright et al., 1988). Furthermore, Albright et al. found significant levels of self-other agreement (r = .23), indicating that people’s ratings of their own attractiveness were related to the perceptions of them by others. However, in both studies, the investigators ignored potential sex differences in their analyses.

The present study used a more elaborate block round-robin design (Kenny, 1990). Every member of each group rated every other member, but each group consisted of two separate blocks: 4 male and 4 female participants. This design resulted in four different variance partitionings—two involving round-robins (the men in the group rating the other men and the women rating the other women) and two resulting from block designs (in which the men rated the women and the women rated the men). By comparing the amount of target variance in each of the four conditions (i.e., men rating men, men rating women, women rating men, and women rating women), we were able to determine whether the degree of consensus in judgments of attrac-
tiveness varies as a function of the sex of the perceiver and the target.

This design also permitted us to examine the related question of whether men and women form similar judgments when rating the attractiveness of others. In other words, are women who are attractive to men also judged as attractive by other women? Are men who are attractive to women also judged as attractive by men? Although many of the features identified by attractiveness researchers appear to apply to both men and women (e.g., “average” faces and symmetry), each sex’s usual interest in the other sex may allow more idiosyncrasy in judgments of members of one’s own sex. In particular, given the culture’s relative focus on female attractiveness (e.g., Udry & Eckland, 1984), there may be less agreement between the sexes in their judgments of men than in their evaluations of women. We were able to address this question by examining target-target correlations, correlating the target effects resulting from men’s perceptions of women with the target effects resulting from women’s judgments of the same female targets and, conversely, correlating the target effects produced by the male and female perceivers when they rated men.

**Meta-Perceptions of Attractiveness and Self-Other Agreement**

We also examined men’s and women’s meta-perceptions of how attractive they are to others. Such judgments are important, affecting people’s thinking about who is likely to return their interest and whom they should reasonably approach (Shanteau & Nagy, 1979), but they may be grounded more in stereotype and wishful thinking than in factual reality (Ashmore & Longo, 1995). For instance, women overestimate the extent to which men seek blonde hair, blue eyes, and large breasts in a woman, and men exaggerate women’s preferences for physical size, blonde hair, and blue eyes in a man (Jacobi & Cash, 1994). Men, although not women, also may tend to believe that they are personally more attractive to the other sex than they really are (Gabriel, Critelli, & Ee, 1994).

We examined the global accuracy with which men and women estimated others’ judgments of them; in addition to rating others, the participants in our study rated themselves and indicated their beliefs about how the others rated them. The SRM also can be used to partition these meta-perceptions into target, perceiver, and relationship variance; in this case, perceiver variance indicates the degree to which respondents believe that everyone rates them similarly, and target variance indicates the degree to which respondents agree that particular judges are more lenient and others more critical. Furthermore, once these meta-perceptions have been partitioned, it is possible to address the question of meta-accuracy (Kenny & DePaulo, 1993): Do people know how they are typically seen by others? Generalized meta-accuracy is assessed by correlating people’s target effects (how they are seen by others) with their perceiver effects for meta-perceptions (how they believe others see them). Examining self-other agreement raises a related but slightly different question: Are people’s self-ratings consistent with how they are seen by others? This question can be addressed by correlating people’s self-ratings with their target effects. Both the meta-accuracy and the self-other correlations may vary as a function of the perceiver and target sex. For example, women’s self-images may be more closely related to how they are perceived by men than by other women. In addition, men may be better able to predict how they are perceived by women than by other men.

**The Present Study**

In groups of eight, four men and four women who had never met evaluated the physical attractiveness of each of the other people in their group and were judged by each of them in turn. We thus provided participants with multiple targets of both sexes in a live, face-to-face setting—the 1st day of a college class—in which naturalistic judgments like these are often made. To maximize the ecological validity of the study, the instructions to the participants were intentionally open-ended: They were not asked to focus on particular features (e.g., face, body shape) or to apply specific standards when judging the attractiveness of their classmates. Unlike studies in which the experimenters selected the physical features of interest (e.g., by providing silhouettes that manipulated body shape or by using photographs that showed only faces), the participants in the present study were free to attend to whatever features were salient to them when making their judgments of attractiveness. For example, some of the participants may have focused on facial features, whereas others attended to body shape. Thus, any consensus in attractiveness judgments would not indicate that perceivers necessarily agree about which features are attractive but about which people are generally attractive.

We also asked participants to rate their own looks and to estimate how they were being rated by each of the other people in their group. The block round-robin design allowed us to determine whether (a) consensus, (b) assimilation, (c) self-other agreement, (d) meta-perceptions, and (e) meta-accuracy were moderated by sex of the perceiver and the sex of the target. Finally, we obtained self-reports of self-esteem that allowed us to explore the links, if any, between self-esteem and one’s actual and perceived attractiveness to others.
METHOD

Participants

The participants were 112 male and 112 female volunteers recruited from several sections of an Introduction to Psychology course at Sam Houston State University, a public university of intermediate size (enrollment 12,000) in the Southwestern United States.

Procedure and Materials

The students received a brief description of the procedure at the end of the first class meeting of the semester, and those who wished to participate were instructed to form groups consisting of four men and four women in which they knew no one else. Experimenters helped arrange participants until the groups consisted only of strangers. In all, 28 different 8-person groups participated. Because each group had to consist of exactly 4 men and 4 women, all of whom were unacquainted, students who did not wish to participate had ample opportunity to decline participation without drawing attention to themselves. Few students declined participation, and there were students in each class who wished to participate but were given an alternate task because they could not be fit into any of the groups.

Once the groups were formed, a code letter ranging from A to H was placed on each person’s desk to identify him or her. In half of the groups, women were given the letters A-D and men the letters E-H; this pattern was reversed in the remaining groups, and in all cases assignment of letters to participants within each block was random. Then, after they provided their informed consent, participants completed three questionnaires.

The first asked participants to take a good look at each of the other members of their group and then to rate the physical attractiveness of each member, in alphabetical order by code letter, on a 7-point Likert-type scale (with 1 indicating extremely unattractive and 7 indicating extremely attractive). As part of this process, participants also provided self-reports of their own physical attractiveness, rating themselves as they did others. To minimize the awkwardness some participants may have felt at rating the attractiveness of same-sex peers, instructions emphasized that these ratings should reflect how objectively attractive the targets were and not whether the respondent was personally attracted to the target. Participants were assured that their ratings were confidential and would not be shared with their fellow group members.

The second questionnaire assessed meta-perceptions of attractiveness and asked participants to indicate, using the same 7-point scale, what rating they believed each member of their group had given them. Finally, on a third form, participants completed three subscales (the “self-regard,” “social confidence,” and “physical appearance” measures) from Fleming and Courtney’s (1984) Self-Rating Scale, a valid and reliable measure of self-esteem. To simplify subsequent analyses, these three intercorrelated subscales (average $r = .60$) were combined to yield a single self-esteem score. This 24-item composite measure was internally consistent, with $\alpha = .93$. Items were scored so that higher scores indicated higher self-esteem.

Once these data were collected, the participants were debriefed and thanked. Kenny’s BLOCKO (1995) and SOREMO (1993) programs, which perform social relations analyses on data collected from block and round-robin designs, respectively, were used for data analyses.

RESULTS

Preliminary Analyses

Prior to performing the variance partitioning, we examined the means and variances of the attractiveness ratings across the four conditions. To examine group differences in these ratings, we used a mixed-design two-way ANOVA with sex of the perceiver as a between-subjects factor and sex of the target as a within-subjects factor. To avoid violating the assumption that our observations were independent, the groups were the units of analysis. As indicated in Table 1, there was a main effect for sex of the target, with women being rated as more attractive than men by both male and female raters, $F(1, 54) = 29.92, p < .001, MSE = .20$.

Because each participant provided multiple ratings on the attractiveness scale (one for each group member), the standard deviation would not be an appropriate measure of variance. Instead, we examined the total variance yielded by these round-robin attractiveness ratings. To examine whether the total amount of variance varied as a function of the sex of the perceiver and the sex of the target, we conducted another mixed-design ANOVA with the group as the unit of analysis, but this time with the total variance in each condition as the dependent variable. As indicated in Table 1, there was a main effect for sex of the perceiver, with the ratings provided by the male perceivers having greater total variability than the ratings of the female perceivers. However, it was necessary to examine the decomposed variances to determine the cause of this greater variability.

Judgments of Male and Female Attractiveness

The first set of SOREMO analyses partitioned the dyadic ratings of attractiveness into target components, perceiver components, and relationship/error components. The relative variance partitioning for these ratings is provided in Table 2 (along with the variance partitioning for the meta-perceptions that will be discussed
below). These values were calculated by dividing the absolute amount of variance that could be attributed to each component by the total amount of variance in the ratings. Thus, they indicate what percentage of each rating could be attributed to target influences and what percentage was due to perceiver characteristics. Relationship effects are not included in the table because they were not subjected to significance testing; participants’ judgments were only made once, and any relationship effects were confounded with error (Kenny & La Voie, 1984).

Consensus. As Table 2 shows, there was meaningful consensus among perceivers (i.e., significant target effects) for judgments of others’ attractiveness regardless of whether women rated women, women rated men, men rated women, or men rated men. However, the degree of consensus appeared to vary with the sex of the person making the judgment and that of the person being judged. Men substantially agreed with one another about the attractiveness of women, for instance, with target effects accounting for 41% of the variance in their judgments. However, there seemed to be less similarity in men’s perceptions when they rated the attractiveness of other men; significant consensus occurred, but it accounted for only 20% of the variance in same-sex male judgments. In contrast, women agreed with each other to much the same degree when they were evaluating the looks of either men or women.

We examined the consensus in these various judgments with a mixed-design two-way ANOVA that used sex of the perceiver as a between-subjects factor and sex of the target as a within-subjects factor (see Table 3). Because the relative target variance depends on the size of the other variance components (i.e., perceiver, relationship/error), the dependent variable for this ANOVA was the absolute target variance. The units of analysis were the 8-person groups of participants. There were no main effects of either the sex of the perceiver, F(1, 54) = 2.42, ns, or the sex of the target, F(1, 54) = 2.69, ns, but there was a significant interaction of the two variables, F(1, 54) = 4.04, p < .05, MSE = .30. Tests of simple effects indicated that this interaction resulted primarily from the high level of consensus among men who were judging the attractiveness of women. Men agreed more in their ratings of women than they did in their ratings of other men, F(1, 54) = 6.66, p < .05. There was also greater consensus in men’s ratings of women than there was in women’s ratings of the same female targets, F(1, 54) = 4.61, p < .05, MSE = .35. The target variance for women’s ratings of men did not differ significantly either from men’s ratings of those same men or from the women’s ratings of other women. Thus, there seemed to be especially high consensus among men about what was and was not attractive in women, and this agreement was higher than either sex’s consensus about what was attractive in men.

Assimilation. There was also noticeable assimilation (i.e., significant perceiver effects) apparent in all four judgments (see Table 2). Although diverse observers tended to agree in their perceptions of others’ attractiveness, such judgments were also in the eye of the beholder to a significant extent, regardless of who was judging whom. There was obvious idiosyncrasy in individuals’ perceptions of strangers’ attractiveness.

<table>
<thead>
<tr>
<th>TABLE 1:</th>
<th>Mean Attractiveness Ratings and Total Variance in Ratings as a Function of the Sex of the Target and Respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex of Target</td>
<td>Female</td>
</tr>
<tr>
<td>Mean attractiveness</td>
<td></td>
</tr>
<tr>
<td>Sex of Respondent</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4.54 (.49)</td>
</tr>
<tr>
<td>Male</td>
<td>4.42 (.33)</td>
</tr>
<tr>
<td>Total</td>
<td>4.48</td>
</tr>
<tr>
<td>Total variance</td>
<td></td>
</tr>
<tr>
<td>Sex of Respondent</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.17 (.80)</td>
</tr>
<tr>
<td>Male</td>
<td>1.63 (.83)</td>
</tr>
<tr>
<td>Total</td>
<td>1.40</td>
</tr>
</tbody>
</table>

NOTE: Standard deviations (with group as the unit of analysis) are provided in parentheses. Within each analysis, values with the same single-letter subscript are significantly different, p < .05. Female and male refer to the sex of the respondent.

<table>
<thead>
<tr>
<th>TABLE 2:</th>
<th>Relative Variance Partitioning for Perceptions and Meta-perceptions of Attractiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex of Target</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>Perceiver Variance</td>
</tr>
<tr>
<td>Attractiveness judgments</td>
<td></td>
</tr>
<tr>
<td>Female perceivers</td>
<td>.35*</td>
</tr>
<tr>
<td>Male perceivers</td>
<td>.19*</td>
</tr>
<tr>
<td>Meta-perceptions</td>
<td></td>
</tr>
<tr>
<td>Female perceivers</td>
<td>.65*</td>
</tr>
<tr>
<td>Male perceivers</td>
<td>.71*</td>
</tr>
</tbody>
</table>

NOTE: The combined relationship and error variance was not submitted to significance tests and has been omitted for purposes of clarity. Nevertheless, it can be readily calculated from the tabled data; it is the percentage of variance remaining for each judgment once perceiver and target variance has been explained. *p < .05.
than in their judgments of women, \(F(1, 54) = 7.08, p < .01\), or in women’s judgments of men, \(F(1, 54) = 9.68, p < .01\), \(MSE = .54\). The perceiver effects for women’s ratings of other women did not differ significantly from their ratings of male targets or from the men’s ratings of the same women.

Consensus versus assimilation. The relative size of the variance components can be concisely described by the ratio of the perceiver variance to the target variance (Malloy et al., 1995). When this ratio is obviously greater than one, perceiver influences account for more of the variance than do target influences, and perceiver idiosyncrasy is the more important factor in interpersonal judgments. In contrast, when the ratio is less than one, there is greater consensus than assimilation, and target characteristics play a larger role in interpersonal perception. It is noteworthy that when both men and women were asked to rate members of the other sex, target effects accounted for about twice as much of the variance as did the perceiver effects (for men’s judgments of women, \(P/T = .46\), and for women’s judgments of men, \(P/T = .45\)). Conversely, perceiver influences outweighed the target influences in same-sex judgments, especially when the men rated the attractiveness of other men (\(P/T = 2.70\)).

Target-target correlations. Did men and women make similar judgments when rating the male and female targets? A high correlation between the target effects that were produced when the men rated the female targets and the target effects yielded by the women when they rated the same female targets indicated that women who were attractive to men were also judged to be attractive by women (\(r = .75, p < .001\)). Similarly, men who were seen as attractive by women were also rated as attractive by men (\(r = .61, p < .001\)). Although idiosyncrasy was apparent in same-sex judgments, men and women generally agreed about who was, and who was not, attractive, both in their own and in the other sex.

Self-Other Agreement and Self-Esteem

Further analyses examined the correlations of the target effects obtained above with the participants’ reports of their own attractiveness and self-esteem. These correlations assessed whether (a) people who were considered attractive by others actually thought themselves to be good-looking—that is, whether there was self-other agreement in these judgments—and (b) whether people who were seen as attractive reported higher self-esteem. It should be noted that these correlations (as well as the following correlations involving variance components) were half disattenuated to take into account the reliability of each variance component (Kenny, 1994); as a result, on occasion, the level of statistical significance (derived from the raw correlation) may not appear to correspond to the magnitude of the disattenuated correlation.

As Table 4 shows, there was significant self-other agreement for ratings of attractiveness. In particular, the men’s judgments of their own looks were closely related to the judgments of them by both male (\(r = .52\)) and female (\(r = .53\)) observers. Men who were considered good-looking by others thought themselves attractive as well. On the other hand, there were no significant links between individuals’ self-esteem and how attractive others perceived them to be.

Meta-Perceptions of Attractiveness and Meta-Accuracy

Table 2 also illustrates the variance partitioning for meta-perceptions of attractiveness. There were no significant target influences, a finding that replicates previous meta-perception studies (Kenny & DePaulo, 1993) and demonstrates that respondents did not discriminate among judges; they did not believe that some judges would find them attractive, whereas others would not. On the other hand, there were substantial perceiver variances, indicating that some respondents believed that they were consistently considered attractive by most people, whereas others believed that they were routinely perceived to be less attractive.

Correlational analyses revealed that some of these perceiver influences were related to the respondents’ self-ratings of attractiveness. As the first row of correlations in Table 5 indicates, women who considered themselves attractive believed that both men (\(r = .71\)) and women (\(r = .72\)) found them to be attractive as well. Men who considered themselves good-looking believed that women also found them attractive (\(r = .40\)), but they did not necessarily assume that other men found them attractive (\(r = .17\)).

Not surprisingly, several of these perceiver effects also were related to the participants’ self-reports of self-


TABLE 4: Self-Other Agreement Correlations for Attractiveness and the Correlations Between Self-Esteem and the Target Effects for Perceptions of Attractiveness

<table>
<thead>
<tr>
<th>Perceiver Effects for Meta-Perceptions of Attractiveness</th>
<th>Perceptions by</th>
<th>Perceptions by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female Perceivers</td>
<td>Male Perceivers</td>
</tr>
<tr>
<td>Self-rated attractiveness of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female targets</td>
<td>.28*</td>
<td>.29*</td>
</tr>
<tr>
<td>Male targets</td>
<td>.53***</td>
<td>.52***</td>
</tr>
<tr>
<td>Self-esteem of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female targets</td>
<td>.23</td>
<td>.12</td>
</tr>
<tr>
<td>Male targets</td>
<td>.25</td>
<td>.21</td>
</tr>
</tbody>
</table>

*p < .05. ***p < .001.

TABLE 5: Self-Rated Attractiveness and Self-Esteem Correlated With Perceiver Effects for Meta-Perceptions of Attractiveness and Meta-Accuracy Correlations

<table>
<thead>
<tr>
<th>Perceiver Effects for Meta-Perceptions of Attractiveness</th>
<th>Perceptions by</th>
<th>Perceptions by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female Perceivers</td>
<td>Male Perceivers</td>
</tr>
<tr>
<td>Self-ratings of attractiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.72***</td>
<td>.71***</td>
</tr>
<tr>
<td>Men</td>
<td>.40***</td>
<td>.17</td>
</tr>
<tr>
<td>Self-esteem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.32**</td>
<td>.35**</td>
</tr>
<tr>
<td>Men</td>
<td>.39**</td>
<td>.15</td>
</tr>
<tr>
<td>Target effects for perceptions of attractiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.26*</td>
<td>.39*</td>
</tr>
<tr>
<td>Men</td>
<td>.49**</td>
<td>.31*</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.


estem. Generally, people who reported higher self-esteem believed that others thought them to be physically attractive, with one exception: Men’s self-esteem was not significantly related \( (r = .15) \) to their beliefs about how attractive they were to other men. Thus, neither men’s perceptions of their own attractiveness nor their self-esteem were related to their beliefs about how physically attractive other men found them to be.

These findings raise the issue of generalized meta-accuracy: Were the people who believed that others saw them as attractive correct? This issue was addressed by correlating the target effects for ratings of attractiveness with the perceiver effects for meta-perceptions of attractiveness. As can be seen from the third set of correlations presented in Table 5, there was a considerable degree of generalized meta-accuracy. In particular, women seemed to know how they were judged by men \( (r = .49) \). Furthermore, even though men who rated themselves as attractive did not necessarily believe that they would be seen as attractive by other men, the meta-accuracy for male targets rated by male perceivers \( (r = .31) \) was similar to the levels of meta-accuracy found in the other conditions.

DISCUSSION

The present study examined young adults’ perceptions of the physical attractiveness of their peers. It differed from most other attractiveness studies in three crucial respects. First, it was designed specifically to examine the ways in which perceiver and target sex may moderate consensus in attractiveness judgments. Second, participants did not form their judgments using limited or static information about their targets; instead, they rated classmates who were physically present, a few feet away. Moreover, the perceivers also served as targets whose own attractiveness was judged. Our procedure thus simulated typical real-life situations in which peers simultaneously evaluate one another, and it had greater ecological validity than many prior studies of attractiveness. Finally, the present study examined the relationships among people’s self-perceptions and evaluations of them by others, and it determined whether these relationships are moderated by the sex of those involved.

The Composition of Attractiveness Judgments

Consistent with prior studies (see Feingold, 1990; Langlois et al., 2000, for reviews), there were significant and substantial levels of consensus for judgments of attractiveness regardless of the sex of the target or the sex of the perceiver. Faced with multiple targets, diverse observers generally did agree about who is, and who is not, pretty or handsome.1

Despite significant consensus across all conditions, the sex of the perceiver and target clearly influenced these judgments of attractiveness. In particular, there was higher consensus when men rated the attractiveness of women than in any of the other three conditions, and there was greater idiosyncrasy (i.e., perceiver variance) when men rated the attractiveness of other men. Indeed, the dissimilarity of these judgments was evident in their ratios of perceiver/target variance: There was more than twice as much target variance as perceiver variance when men rated the attractiveness of women but more than twice as much perceiver variance as target variance when men rated other men. Thus, men appeared to engage in very different processes when they judged women as opposed to men; they carefully discriminated between attractive and unattractive women, agreeing among themselves, but displayed more diversity and idiosyncrasy in their ratings of other men.2
In contrast, the attractiveness ratings made by women were not substantially influenced by the sex of the target; female perceivers agreed among themselves to a similar extent whether they rated women (29% target variance) or men (31% target variance). Notably, the composition of women’s judgments of the other sex was very similar to that of men; the ratio of the perceiver/target variances was .45 when women evaluated men and .46 when men judged women. There appeared to be more perceiver variance in women’s ratings of other women than in their ratings of men, but the difference between the two proportions was not statistically significant.

Overall, these differing levels of consensus are consistent with theories that emphasize mechanisms of mate selection in the evolution of attractiveness preferences (e.g., Buss, 1999; Cunningham, 1986). Both men and women agreed among themselves about the attractiveness of members of the other sex. However, there are presumably greater evolutionary pressures on men than on women to be judicious about the youth and health (and thereby fertility) of potential partners, and careful assessment of the physical attractiveness of women should result. Indeed, men were more consistent in their judgments of the other sex than were women. A contrasting sociocultural perspective would reasonably predict that both men and women should display notable consensus in their judgments of women, and such agreement did occur. However, proponents of a sociocultural perspective still might argue that there are social forces that contribute to heterosexual men being more likely than women to agree among themselves when evaluating women.

These data also shed light on the self-perceptions of discernment and perspicacity in attractiveness judgments obtained from men and women by Graziano and his colleagues (Graziano et al., 1993). Women were evidently correct in reporting (in that prior study) that they were better judges than men of the attractiveness of other men, but they were incorrect in assuming that they were as discerning in their judgments of other women. Despite cultural emphasis of female attractiveness, and despite women’s considerable attention to matters of beauty (Kalb, 1999), men agreed even more closely in their ratings of women’s looks than did the women themselves.

Meta-Accuracy and Self-Other Agreement

The participants also provided ratings for two related, but not identical, questions: How attractive did they think they were (self-ratings) and how did they think they were rated by others (meta-perceptions)? Overall, people’s beliefs about their own attractiveness were generally consistent with the manner in which they were seen by others; all four self-other agreement correlations were statistically significant, ranging in magnitude from .28 to .53. In addition, people had some knowledge of how attractive they were to their peers; all four meta-accuracy correlations were also significant, ranging from .26 to .49.

Sex differences in these judgments were less striking than they had been in perceptions of others’ looks, but women did seem to be particularly aware of how they were rated by men; their meta-accuracy (.49) in this instance was the highest of the four combinations. The women’s meta-perceptions also may have been especially influential: Their judgments of their own attractiveness were highly similar to their meta-perceptions of what others thought of them, yielding correlations of .72 and .71. These findings fit the notable fact that 91% of the cosmetic surgery performed in the United States in 1998 was done on women (Kalb, 1999). Interpersonal perceptions of attractiveness appear to carry impressive force in women’s lives in a manner that is consistent with a sociocultural model. In contrast, men’s self-ratings were more modestly related to their judgments of how women rated them (.40) and were not at all related to their judgments of how other men rated them. Compared to women’s self-perceptions, men’s ratings of their own appeal appeared to be less closely tied to evaluations of them by others, a fact that fits the rather generous body images men possess; unlike women, men tend to believe that their physiques are more attractive to others than they really are. Women have poorer body images than do men (Feingold & Mazzella, 1998), and that may be a regrettable cost of being more attuned than men are to social evaluations of their physical attractiveness.

Implications and Limitations

These various results form a pattern that jives nicely with other diverse findings of prior investigations. First, they provide a convincing demonstration that, in this college sample, there are shared standards that allow diverse observers to agree among themselves about the physical attractiveness of the strangers they encounter. (Our procedure did not determine what those standards are, but our results imply that they must exist. Future studies in which objective physical measurements are collected and correlated with the target effects yielded by the SRM will be necessary to determine which standards best account for the attractiveness ratings and whether these standards vary as function of the sex of the perceiver and the target.) Beauty is quite clearly not entirely in the eye of the beholder. Instead, some of us are judged by almost everyone we meet as handsome or lovely, whereas others of us nearly always seem plain. In particular, perceptions of the attractiveness of a member
of the other sex emerge more from consensually held views of that person’s characteristics than from idiosyncratic tastes that are unique to individual perceivers. This consensus probably contributes to the impressive influence of physical attractiveness in social life. When they first meet us, others concur in their assessments of our looks.

Consensus was especially evident in men's ratings of women's beauty, but men's judgments of the attractiveness of other men were more idiosyncratic. Indeed, in general, men's perceptions of the looks of other men operated differently than the other three judgments examined here. We do not know why but we can suggest that the notable idiosyncrasy in these judgments could be due to men either caring little about other men's opinions or caring too much. Because women emphasize status and resources when they evaluate men (Buss & Schmitt, 1993), and because status figures more prominently in men's intrasexual rivalries than do looks (Buss & Dedden, 1990), men may simply be unconcerned with other men's looks. Alternatively, men may defensively avoid any information regarding the attractiveness of other men to avoid giving others the embarrassing impression that they are interested in other men's looks.

It is also possible, of course, that the large amount of perceiver variance in men's ratings of other men is a phenomenon specific to the men in Southeastern Texas, where this study was conducted. A homogeneous, collegiate pool of participants is one of the study's limitations. (However, because nearly all of the students who were given the opportunity to participate did so, we do not believe that we obtained a self-selected sample that differed in any meaningful way from our usual population of psychology students.) Our ratings also were obtained in the absence of any meaningful interaction among the participants; it is likely that the women's—although not the men's—ratings of others may have changed had they discussed the targets with their peers (Graziano et al., 1993).

Future studies in which participants interact before providing attractiveness ratings will be necessary to determine whether consensus varies depending on the participants' level of acquaintance.

On the other hand, the procedure actually possessed more verisimilitude than most other studies of attractiveness, and if one notes that our conclusions may only apply to judgments of attractiveness at first meeting, the results are clear: Men and women agree among themselves regarding who among them is attractive. In particular, although some individuality and idiosyncrasy are clearly present, people's judgments of the attractiveness of members of the other sex exhibit more consensus than assimilation. Furthermore, men and women made similar judgments when rating male and female targets. Participants' self-rated attractiveness also was significantly correlated with their target effects, indicating self-other agreement. In contrast to ratings of attractiveness, meta-perceptions of attractiveness were primarily a function of the perceiver, and there was evidence of generalized meta-accuracy: People who thought that others found them attractive generally received higher ratings. This evidence for generalized meta-accuracy was especially pronounced when women indicated how they believed they were being perceived by men. Overall, we know who is pretty or handsome, and those who are attractive know it as well.

NOTES
1. Although the target variances in the present study appear to reflect lower levels of consensus than those reported in Langlois et al.'s (2000) meta-analysis, this difference reflects the different statistics used. Clearly, there is consensus in judgments of attractiveness, but the very high r values in meta-analysis are effective reliabilities that are inflated by the number of raters. So, for example, the effective reliability of .90 for within-culture agreement was derived from a mean interrater agreement of r = .47. Thus, Langlois et al.'s values should not be interpreted to mean that there is near unanimity in these judgments or that perceiver, relationship, and error variance do not contribute to attractiveness judgments.

2. In a Social Relations Model (SRM) study that included cross-sex judgments of attractiveness, Themanson (1997) found more total variance in attractiveness ratings and higher levels of target variance than we did and failed to find greater consensus when men rated women than when women rated men. However, several methodological differences—for instance, Themanson's participants did not provide same-sex ratings and they rated one another on a series of personality traits in addition to attractiveness—make it difficult to determine why the two studies yielded somewhat inconsistent results.

REFERENCES


