May, 1989

Gender, Sexual Orientation, and Truth-of-Consensus in Studies of Physical Attractiveness

James M Donovan
Elizabeth Hill
William R. Jankowiak

GENDER, SEXUAL ORIENTATION, AND TRUTH-OF-CONSENSUS IN STUDIES OF PHYSICAL ATTRACTIVENESS

JAMES M. DONOVAN,
Tulane University

ELIZABETH HILL, Ph.D.,
University of Michigan

AND WILLIAM R. JANKOWIAK, Ph.D.
Tulane University

Truth-of-consensus methodology presently holds that sex differences in perceptions of physical attractiveness are negligible and may be routinely ignored during prescaling. No determination has been made in the literature of the effects of sexual orientation on this perceptual process. The data presented herein suggest that while sex and sexual orientation of judge are largely irrelevant to prescaling of female stimuli, these variables are important when judging male stimuli. In particular, male homosexuals and male heterosexuals differ significantly in ranking male facial photographs. Thus, experimenters wishing to treat attractiveness levels as known quantities should control for this difference, especially when using a small number of judges for prescaling.

KEY WORDS: gender, physical attractiveness, sexual orientation

Introduction

Physical attractiveness studies depend to a large extent upon the validity of the "truth-of-consensus" methodology. In this method, "If a significant number of 'judges' designate a person as physically attractive, then that person is defined as physically attractive" (Berscheid & Walster, 1974, p. 181). The attractiveness levels of facial photographs subjected to this methodology are thereafter treated as known quantities. Patterning of dependent variables is then attributed to the manipulation of the independent variable of physical attractiveness (e.g., Byrne et al., 1968).

Not everyone in all circumstances is suitable to serve as a judge to determine the physical attractiveness of a stimulus. Some variables, such as race, have been shown to influence perceptions of physical attractiveness (Cross & Cross, 1971). To maximize validity of truth-of-consensus determinations, judges used during prescaling should be matched with the intended subjects for the experiment. Only in this way can the designated level of attractiveness for a given stimulus be treated confidently as an independent variable for manipulation.

Within this context, the influences of subject and stimulus gender on perceptions of physical attractiveness become particularly relevant. According to a recent review, "Research repeatedly reports no significant differences

between the ratings of physical attractiveness given by male and female judges" (Patzer, 1984, p. 20). The context suggests that this conclusion holds regardless of the stimulus gender. An examination of the cited sources, however, shows the matter to be less clear. When Kopera et al. (1971) conclude that "there [are] no significant differences between males and females on the ratings" (p. 317), it should be stressed that the stimuli were all female. Left open is the treatment of male stimuli. Most notably, Cross and Cross (1971) conclude that the interaction between judge and stimulus genders is in fact quite significant.

Recent articles contribute to the debate. Kerr and Kurtz (1978) found sex differences present to the extent that they suggest "attention to the sex of the judges and stimuli is required in the prescaling of physical attractiveness stimuli" (p. 181). Finally, Schulman and Hoskins (1986) report no significant main effects of judge or stimulus gender, nor an interaction between them, on good-lookingness ratings, although they do for other measures.

Given the wide assumption of no gender differences, it would be surprising if it did not hold true at some level. Perhaps limiting the sweeping claim of "no difference" can better explain the contradictory findings. For example, does the lack of differences describe the way the sexes ultimately rank stimuli (the mean), does it pertain to the dispersion of ratings within each gender (the variance), or both? Truth-of-consensus is methodologically dependent upon the comparison of group means (Patzer, 1984), and therefore conclusions of "no difference" should not be expanded to suggest that there are no differences of any kind.

Further, judge gender must also be considered in prescaling and interpretation, even with nonsignificant main effects, especially if it interacts with another variable that is relevant to attractiveness perceptions. For instance, sexual orientation might influence attractiveness ratings, which in turn might interact with judge gender. One, after all, does not encounter "homosexuals" per se, but instead, gay men and lesbians.

This paper reports the findings of a project comparing physical attractiveness judgments by men and women differing in sexual orientation. The original purpose of the project was to determine whether patterns of attractiveness perceptions can best be explained by theories of sex-role socialization or by those from social biology (Donovan et al., 1988; Symons, 1979). However, these data also allowed us to consider whether the truth-of-consensus assumption of no sex differences, based upon a comparison of group means, required revision in light of sexual orientation. It is this post facto analysis of means which is reported herein.

Methods

Stimulus Photographs

Photographs were taken at random from a 1979 yearbook of a major private southern university. The selection process netted photographs of 35 females and 39 males, all Caucasian. Ten judges (5 male, 5 female, including one male homosexual) were asked to rank each of the two photo sets on the criterion of "good looks." The ten ranks for each photo were averaged, and the standard deviations were calculated. Using the averaged rank, the photographs were

renumbered 1-35 (female) and 1-39 (male); if two photographs had an identical mean, the photo with the larger standard deviation was discarded (2 males and 2 females were thus discarded). In order to obtain a manageable group of photographs representing a continuum of "good looks," the photographs were paired (e.g., 1-2, 3-4, etc.), and a member of a pair with the larger standard deviation was discarded. The last (unpaired) photo of each set was retained. This process resulted in a set of 17 females and 19 males; two additional males were haphazardly discarded to establish numerical equality between the two sets.

Additionally, photographs of 6 males and 7 females were taken from the directory of a graduate professional school. After being ranked and averaged according to the process described above, the top three photographs of each sex were selected for inclusion. Rationale for this process was that since the original project goal desired to compare older versus younger stimuli in terms of the rankings each set received, interpretation of between-group differences would be facilitated if the older subset was not also unattractive.

The final number in each set was twenty. Each set is assumed to contain a continuum of attractiveness levels. High quality photostatic copies were made so that identical photo sets could be provided to each of three interviewers. All responses are to copies of the sets; no subject was exposed to the original.

Participants

Caucasian participants were solicited on an opportunistic basis by each interviewer. Respondents were solicited primarily through the authors' friend-ship networks and students; early subjects were encouraged with some success to refer others to be interviewed. Two authors spent one evening in a lesbian bar interviewing subjects.

Although all interviewers (the authors) attempted to solicit data from each of the four study populations (homosexual and heterosexual men and women), interviewers and subjects were not completely counterbalanced. However, examination of data by interviewer suggests that this circumstance did not significantly bias the results. All participants with Kinsey scores of 2-4 (the bisexual middle range) were excluded from analysis because the authors wanted to contrast clear groups in this preliminary investigation. A total of 61 people comprise the final subject pool (17 heterosexual men, 12 homosexual men, 21 heterosexual women, and 11 homosexual women). Homosexual participants were somewhat older than the heterosexuals. However, it appears that subject age is not a critical factor, since Spearman correlations are non-significant between subject age and average rank given the older subset.

Procedure

Initial demographic information was obtained from the subject, including sex, age, and Kinsey score for sex-object orientation. (All responses at the homosexual pole of the orientation scale were presumed to be truthful responses due either to previous acquaintance with the subject or to the setting from which the subject was drawn. Potentially less reliable are claims of heterosexuality by some undergraduates.) The subject was then instructed to rank members of each set of photographs according to how "good-looking" they were. This word was preferred to "beautiful" because the latter frequent-

ly connotes feminine qualities, whereas the former term is comparatively gender-neutral. If needed, clarification was offered, i.e., that this evaluative judgment was to be based on physical appearance alone. Ranks were recorded (1-20) on a score sheet along with the subject's comments. The order of presentation of male and female stimuli was alternated from one subject to the next.

Statistical Analysis

Group differences in mean ranks for photographs were analyzed by Mann-Whitney tests. Main effects for differences between genders and between preferences were tested by collapsing over the other factor. The interaction was tested by analyzing differences between the additive effect of gender for the two preference groups (cf. Bradley, 1968; pp. 138-141). When significant interactions occurred, subsequent comparisons were made between pairs of groups.

Spearman rank correlations were calculated for the 20 photographs using the mean ranks for the four groups. Overall differences among the twelve correlations were analyzed by chi-square after using Fisher's z-transformation (Snedecor & Cochran, 1980, p. 187). This was repeated among the 6 male and 6 female correlations. Pairwise tests were then conducted using Fisher's z-transformation (Snedecor & Cochran, 1980, p. 186). Another measure recorded was the degree of overlap between groups at the extremes of the ranks. Differences from the average ranking were noted.

Results

Several patterns emerge from an examination of the results (Table 1). Perceptions of female stimuli are more consistent across groups than for male stimuli. Including trends (.05 < p < .10) with significant differences, mean ranks differed among groups on only 8/20 female photographs compared to 15/20 male photographs. For female photographs, two were considered more attractive by women and one considered more attractive by men; also, two were considered more attractive by heterosexuals and two more attractive by homosexuals. There were no dramatic interactions.

Male stimuli, however, were frequently given different ranks by different groups, including differences by rater gender, preference, and their interaction. In fact, there were five significant interactions (p < .05). One of these is due to a higher ranking by homosexual women than the other groups (Table 1, photo 6, homo- vs. heterosexual women: U = 175, p = .018). The other four interactions are primarily due to differences between homosexual and heterosexual men. Subsequent pairwise comparison shows significant differences between these two groups on photographs 3 (U = 151, p = .027), 7 (U = 44, p = .01), and 8 (U = 42, p = .008). These groups tended to differ on photo 20 (U = 138, p = .097), but the contrast between homosexual men and women is larger (U = 29, p = .011).

Agreement in the extremes of the distribution is also shown in Table 1. Superscripts indicate preferences disparate from the overall order, based on an average of the four mean ranks. For female stimuli, there was complete agreement on the top five photographs; that is, while specific position varies, the same five photographs are in the top five for all groups. There is also good

agreement for the bottom five. The only exceptions are the rankings of two photographs by homosexual women compared to the other groups. In contrast, there is much less overall agreement in judging male photographs. No two groups agree more completely on membership in the top or bottom five.

This pattern is shown by the Spearman correlations in Table 2. There are overall differences among all 12 correlations ($x^2 = 27.13$; $p < .005$; 11 df). The six intercorrelations among rankings of male photographs do not differ significantly ($x^2 = 3.69$; $p > .05$; 5 df); however, the smallest correlation is between rankings of male photographs by hetero- and homosexual men ($r_s = .589$). This correlation does tend to differ from the highest for male photographs, i.e., between heterosexual and homosexual women ($z = 1.755$, $p = .093$). Similarly, the female intercorrelations do not significantly differ ($x^2 = 5.341$; $p > .05$; 5 df), but the correlation for heterosexual men and women tended to be higher than that for homosexual women and heterosexual men ($z = 1.953$, $p = .051$). On the average, the correlations for female photographs are higher than those for male photographs. In fact, the highest correlation is significantly higher than all but one of the intercorrelations for male photographs ($p < .05$). Also, most of the intercorrelations for female photographs are significantly different from the lowest two correlations for male photographs ($p < .05$).

Discussion

There appears to be more agreement and consistency about the physical attractiveness of women than of men. In the set of photographs studied herein, more male than female photographs were given different ranks by various groups. Similarly, the intercorrelations among ranks by various groups were higher for female than male photographs. Thus, the conclusions of Kopera et al. (1971) are supported to a great extent for female photographs, although there are some significant effects of rater gender and sexual preference. However, for male photographs, there appear to be more differences by rater gender and preference, and interactions do occur between these factors. In particular, the differences between heterosexual and homosexual men suggest potential methodological problems.

For example, where male stimuli are involved, it appears that sexual orientation may play a role during prescaling. Thus, the experimenter should consider controlling for the sexual orientations of potential judges, depending on the probable orientation of the intended subject population. Male stimuli inadvertently prescaled by male homosexuals could be labeled as being highly attractive. If these prescaled attractiveness values are then used to explore various halo effects surrounding physical attractiveness variations, heterosexual males could confound the results because they fail to perceive the stimuli as being physically attractive.

Perhaps if the experimenter used, in Bersheid and Walster's terms, a "significant number of judges" (1974, p. 181), one would not have to consider these additional variables. As indicated above, it is not uncommon for experimenters to define levels of attractiveness based upon the perceptions of a small number of persons, not infrequently limited to the authors themselves (e.g., Kerr & Kurtz, 1978). Subtle effects negligible in larger samples can be greatly enhanced within smaller ones which, when combined with random

variation, could skew the attractiveness level of a stimulus.

The relatively low incidence of homosexuality in the general population should not be construed as insurance against the proposed effect. Subjects for experiments of this sort are not typically drawn randomly from the general population, but instead, from special subgroups (e.g., college campuses generally, and introductory psychology classes specifically). We are aware of no evidence that the percentage of homosexuals within these nonrandom groups is equivalent to that of the general population; indeed, homosexuals might be overrepresented, given the somewhat higher achievement sometimes reported for them (Fay et al., 1989; Weinrich, 1978).

Where larger numbers of judges for prescaling are impractical, and where experimenters are uncomfortable inquiring into sexual orientation, the broad view of these data suggest that male stimuli could be prescaled (in the absence of the suggested precautions) by females only. Female determinations, while different from both homosexual and heterosexual males, appear to be in the middle ground between the two male extremes.

References

BERSCHEID, E., & WALSTER, E. (1974). Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 157-215). New York: Academic Press.

BRADLEY, J. V. (1968). *Distribution-free statistical tests.* Englewood Cliffs, NJ: Prentice-Hall.

BYRNE, D., LONDON, O., & REEVES, K. (1968). The effects of physical attractiveness, sex, and attitude similarity on interpersonal attraction. *Journal of Personality, 36,* 259-271.

CROSS, J. F., & CROSS, J. (1971). Age, sex, race, and the perception of facial beauty. *Developmental Psychology, 5,* 433-439.

DONOVAN, J. M., HILL, E. M., & JANKOWIAK, W. R. (1988). Attractiveness judgements, sex, and sexual orientation. Presented at the Conference on Evolution and Human Behavior, Ann Arbor, Michigan.

FAY, R. E., TURNER, C. F., KLASSEN, A. D., & GAGNON, J. F. (1989). Prevalence and patterns of same-gender sexual contact among men. *Science, 243*(4889), 338-348.

KERR, N. L., & KURTZ, S. T. (1978). Reliability of the "eye of the beholder": Effects of sex of the beholder and sex of the beheld. *Bulletin of the Psychonomic Society, 12,* 179-181.

KOPERA, A. A., MAIER, R. A., & JOHNSON, J. E. (1971). Perception of physical attractiveness: The influence of group interaction and group coaction on ratings of the attractiveness of photographs of women. *Proceedings of the 79th Annual Convention of the American Psychological Association, 6,* 317-318.

PATZER, G. L. (1984). *The physical attractiveness phenomena.* New York: Plenum Press.

SCHULMAN, G. I., & HOSKINS, M. (1986). Perceiving the male versus the female face. *Psychology of Women Quarterly, 10,* 141-154.

SNEDECOR, G. W., & COCHRAN, W. G. (1980). *Statistical methods (7th Ed.).* Ames, IA: The Iowa State University Press.

SYMONS, D. (1979). *The evolution of human sexuality.* New York: Oxford University Press.

WEINRICH, J. D. (1978). Nonreproduction, homosexuality, transsexualism, and intelligence: I. A systematic literature search. *Journal of Homosexuality, 3,* 275-289.

Table 1

Mean ranks given male and female photographs (He = heterosexual; n = 17 men, 21 women; Ho = homosexual; n = 12 men, 11 women).

| | Photographs of Men | | | | | | |
| | Male | | Female | | Significant Effects | | |
Judge:	He	Ho	He	Ho	Gender	Pref[b]	G*P
Order:							
1	1.6	2.3	2.3	3.1	*	—	—
2	6.6	4.6	6.3	7.0	—	—	—
3	8.0[a]	4.6	6.1	7.8	—	—	*
4	7.5[a]	6.1	8.4	6.0	—	+	—
5	6.8	6.8[a]	6.2	9.8[a]	—	+	—
6	7.4[a]	8.2	9.3	5.1[a]	—	—	*
7	6.1[a]	10.8	8.2	8.2	—	*	*
8	8.6	13.6[a]	9.3	8.8	—	+	**
9	9.5	10.3	9.2	10.1	—	—	—
10	7.1	11.3	9.2	11.1	—	*	—
11	11.6	12.0	8.0	9.5	*	—	—
12	12.6	10.8	10.0	10.3	—	—	—
13	13.2[a]	6.7[a]	12.6	9.0	—	**	+
14	12.4	11.8	12.8	10.5	—	—	—
15	14.6[a]	9.6	14.3[a]	12.4	—	**	+
16	11.2[a]	15.4	11.6[a]	14.5	—	**	—
17	12.8[a]	13.0[a]	13.3	16.5	+	—	—
18	15.6	14.8	15.3	14.8	—	—	—
19	17.4	19.3	18.2	15.9	*	—	—
20	19.1	18.1	19.1	19.7	+	—	*

| | Photographs of Women | | | | | | |
| | Male | | Female | | Significant Effects | | |
Judge:	He	Ho	He	Ho	Gender	Pref[b]	G*P
Order:							
1	2.0	4.6	3.2	7.1	*	**	—
2	4.2	4.3	5.4	3.8	—	—	—
3	5.7	5.3	3.6	2.9	*	—	—
4	6.1	5.6	4.6	7.1	—	—	+
5	6.2	6.0	6.5	6.6	—	—	—
6	8.5	6.3	8.5	9.5	—	—	+
7	8.5	7.8	10.1	9.5	—	—	—
8	8.8	9.6	8.6	8.9	—	—	—
9	8.8	10.7	6.7	9.0	—	—	—
10	9.8	9.4	10.8	11.6	—	—	—
11	9.7	9.5	11.0	12.5	—	—	—
12	12.9	10.7	12.2	11.0	—	—	—
13	10.6	14.1	12.5	12.2	—	—	—
14	14.4	14.6	12.1	11.9	**	—	—
15	10.6	13.9	12.6	14.6[a]	—	**	—
16	15.7	15.8	15.4	13.8	—	—	—
17	15.3	16.3	14.6	15.6	—	—	—
18	17.5	14.8	16.7	12.1[a]	—	**	—

Table 1—continued

19	17.5	14.8	17.3	15.0	—	*	—
20	17.2	16.5	17.4	15.2	—	—	—

[a]For the highest and lowest five ranks, cases in which a group placed a photograph differently than the overall average of all four groups.
[b]Sexual preference.
$+.05 < p < .10.$ $*p < .05.$ $**p < .01.$

Table 2

Matrix of Spearman Correlation Coefficients

	Males		Females		
	Hetero-sexual	Homo-sexual	Hetero-sexual	Homo-sexual	
Males					
Heterosexual	1.00	.59	.88	.77	Male Photos
Homosexual	.95	1.00	.70	.77	
Females					
Heterosexual	.96	.93	1.00	.80	
Homosexual	.87	.89	.92	1.00	
	Female Photos				