Exploring Stimulus Variability in Applicant Attractiveness

8-13-2013

Robert L. Dipboye
University of Central Florida, rdipboye@knights.ucf.edu

Lyndsey Dhahani
University of Central Florida

Original Citation

Dipboye, R. L. & Dhahani, L. (August 13, 2013). The magnification, mitigation, and reversal of the physical attractiveness effect. Annual Meetings of the Academy of Management, Orlando, Fl

Exploring the Effects of Physical Attractiveness in Job Applicant Evaluations: Taking Into Account Stimulus Variability

Robert L. Dipboye and Lyndsey Dhahani

University of Central Florida

dipboye@rice.edu

This is a version of a paper ("The magnification, mitigation, and reversal of the physical attractiveness effect") presented at the August, 2013, Annual Meetings of the Academy of Management, Orlando, Fl.

Abstract

Previous research on physical attractiveness bias in job applicant evaluations has ignored three important issues. First, the sex-typing of the positions for which applicants are evaluated is usually weak despite the need to provide strongly male and female-typed positions in testing for beauty is beastly effects. Second, the samples of stimuli used in the manipulations of applicant sex, attractiveness, and sex-typing of the job are small. Third, the statistical analyses used in testing hypotheses fail to incorporate variability among both human participants and stimuli. The present research corrected for these three omissions in an experiment in which participants evaluated the suitability of applicants who were physically attractive or unattractive, male or female, and were applying for a male-typed or female-typed position. The experimental design was a within-person 2 (applicant sex) X 2 (applicant attractiveness) X 2 (sex-type of job) ANOVA. Each participant received a set of eight applicants with the photograph used in the manipulation of sex and attractiveness and the type of job randomly drawn from a pool of photographs and jobs. Consistent with the recommendations of Clark (1973), the hypotheses were tested using as subjects the human participants (F1 analyses), pictures (F2 analyses), positions (F2 analyses), and picture-position combinations (F2 analyses). Also, quasi-Fs were conducted to incorporate variability of both human participants and stimuli. All the analyses revealed an attractiveness bias in which the attractive candidates were evaluated more favorably than unattractive candidates. A job sex-type X applicant sex interaction revealed that males were evaluated more favorably for male-typed positions and females for female-typed positions. Also found were main effects for applicant sex and job-type, although these effects were subsumed by the job sex-type X applicant sex interaction. The analyses of the ratings of suitability were consistent with chi-square analyses of best- and worst-fit choices. The findings suggest that the bias against attractive applicants is robust across stimuli as well as human participants. No

evidence was found for a beauty is beastly effect. Exploratory analyses suggested that a bias against attractive females is limited to a narrow domain of jobs.

Exploring the Effects of Physical Attractiveness in Job Applicant Evaluations: Taking Into Account Stimulus Variability

Researchers have reported a bias against people who are relatively unattractive in their physical features and a bias in favor of those who are relatively attractive in a variety of domains (Eagly, Ashmore, Makhijani & Longo, 1991; Hosada, Stone-Romero & Coats, 2003). Moreover, beauty is not in the eye of the individual beholder. Raters across diverse cultures appear to agree on what constitutes facial attractiveness (Langlois, Kalakanis, Rubenstein, Larson, Hallam, & Smoot, 2000) with the features of an attractive face including symmetry, averageness, and sexual dimorphism (Rhodes, 2006). The cognitive interpretation of this "beauty is good" effect states that the bias against unattractive persons is mediated by attributions of trustworthiness, competence, social skill, and a variety of other positive traits (Eagly, et al, 1991; Feingold, 1992; Jackson, Hunter, & Hodge, 1995; Langlois, et al, 2000). Another interpretation is that the bias in favor of the attractive is rooted in unconscioius, affective responses that are hardwired as a consequence of human evolution and associated with physical characteristics associated with successful reproduction (Bzdok, Langner, Caspers, Kurth, Habel, Zilles, & Eickhoff, (2011). One can extrapolate from this interpretation to propose that the bias against the physically unattractive is so deeply rooted that it generalizes beyond mate selection to a variety of domains including personnel selection.

Whether mediated by trait inferences or affective responses, the end result is to place the physically attractive person at an advantage over relatively unattractive persons. This discrimination has been demonstrated across a variety of domains. The physically attractive are evaluated more favorably as romantic partners (Feingold, 1990; Eastwick, Luchies, Finkel, & Hunt, 2013), as defendants in court trials (Mazzella & Feingold, 1994), as students by teachers

(Ritts, Patterson, & Tubbs, 1992), and even as political candidates (Bnaducci, Karp, Thrasher & Rallings, 2008). The results of meta-analyses of the physical attractiveness research are consistent in showing support for a beauty is good effect (Eagly, Ashmore, Makhijani, & Longo, 1991; Eastwick et al, 2013; Hosoda, et al, 2003; Langlois et al, 2000). We are most concerned in the present study with the research showing that physically attractive job applicants are evaluated as more suitable for hiring than physically unattractive applicants. In a meta-analysis of this research, Hosada et al (2003) found a substantial bias against unattractive persons (d = .37). Although the effect had declined somewhat over time and was stronger in within-subjects designs than between-subjects designs, the bias occurred regardless of the amount of information on the applicant, the student vs nonstudent status of the rater, and the gender of the rater.

Despite the support for a beauty is good effect in the evaluation of job applicants, some researchers have concluded that the typical attractiveness bias is moderated by the sex of the applicant and gender typing of the position. Specifically, a beauty is beastly effect is reported in which attractiveness is a disadvantage for women applying for traditionally male positions. In a widely cited laboratory experiment, Heilman and Saruwatari (1979) conducted a 2 X 2 X 2 between groups design in which college students were presented with an attractive or unattractive, male or female applicant. They evaluated the suitability of the applicant for either a female-typed job (secretary) or a masculine-typed job (management trainee). When the position was a nonmanagerial position (secretary), the attractive candidate was evaluated more favorably than the unattractive candidate, regardless of the sex of the applicant. When the position was a managerial position, the attractive male candidate was evaluated more favorably than the unattractive male candidate, but this effect was reversed for the female applicant. In other words,

beauty was beastly in that the attractive female applicant was apparently denigrated relative to the unattractive female applicant. The authors explain this effect with a person-job fit model that emphasizes the cognitive mediators of attractiveness bias. They hypothesize that attractiveness has the effect of enhancing the perceived femininity of the attributes of the female candidate. As a consequence, the attractive female applicant is seen as providing a poorer fit to the masculine-typed managerial position than the unattractive female applicant.

Despite the growing volume of research on the effects of physical attractiveness on evaluations of job applicants, there are three limitations in the previous research that cast some doubt on the robustness of attractiveness biases in evaluations of job applicants. One issue concerns the gender-typing of the male and female positions used in examining the effects of applicant sex and attractiveness. To determine whether attractiveness benefits or harms an applicant, it is important to have participants rate applicants on their qualifications for strongly sex-typed male and female positions. Based on the job-fit model that Heilman et al (1979) proposed in the original demonstration of the beauty is beastly effect, a bias against attractive female applicants occurs when the position is strongly typed as a male-typed job requiring masculine traits and occupied mostly by men. The beauty is good effect occurs when the position is a female-typed job requiring feminine traits and occupied mostly by women. Notwithstanding the importance of the gender-typing of the position, few studies provide a manipulation of the sex-type of the position, with many using moderately sex-typed or gender neutral positions or not even specifying the nature of the position.

The second issue concerns the sampling of the stimuli used in depicting the applicant and the position. Researchers typically have used very few stimuli to manipulate attractiveness and gender of the applicant. Although the procedures used in selecting the photographs used in the

manipulation of sex and attractiveness are usually rigorous, only a few photographs are typically used to represent the attractive and unattractive, male and female applicants. Stimulus sampling is even more of a problem with regard to the positions for which applicants are evaluated. When there is a manipulation of job sex-typing, only one position typically is used to reflect the male-typed and female-typed positions. The use of narrow stimulus samples poses several potential threats to the validity of the interpretations of experimental findings. First, narrow samples risk confounding the manipulation with factors associated with the stimuli used to represent experimental conditions. If the one or two photographs used to represent the unattractive condition depicted a person dressed inappropriately for an interview whereas the photographs representing the attractive condition depicted more appropriate clothing, appropriateness of the dress is a confound that may account for effects of the manipulation rather than facial attractiveness. Second, narrow samples pose a threat to the generalizability of the findings across a broader range of stimuli representing the manipulated variable. For instance, if the attractive female condition was depicted with a glamour shot of a blonde, Marilyn Monroe type applicant, one might not be able to generalize the effects to other types of female attractiveness (e.g., a highly competent looking but attractive applicant).

The third issue concerns the statistical analyses conducted in previous research on attractiveness bias. Even when more than one stimulus is sampled, none of the previous research has incorporated stimulus variability in the statistical tests. When more than one stimulus is used to depict each condition, the evaluations of each stimulus are aggregated in assessing the statistical significance of differences between conditions. In these analyses, the aggregated rating of the stimuli is used as a fixed-effect rather than including variability in ratings among stimuli and treating the manipulation as a random effect. The effects of variables manipulated with

stimuli such as photographs and job titles may prove less robust when stimulus variability is incorporated in the statistical tests of the hypotheses.

The issue of stimulus sampling has been raised several times as a potential problem in psychological research. Clark (1973) originally drew attention to this issue in a criticism of verbal learning research. He noted that limited samples of words were typically drawn and results were analyzed using a fixed effects ANOVA. He recommended the use of larger word samples and the use of analyses that take into account variation across stimuli as well as participants. The failure to account for stimulus variability in statistical tests of hypotheses is also a potential problem in experimental research in both social psychology (Judd, Westfall & Kenny, 2012) and industrial and organizational psychology (Fontenelle, Phillips & Lane, 1985; Highhouse, 2009). Typically, a fixed effects analysis of variance is used to test for the statistical significance of the effects of the manipulations (referred to as F1). Despite the variation in stimuli sampled in creating each condition (e.g., attractive female applying for a managerial job), the different stimuli are considered equivalent. In other words, a fixed effects model assumes that variability in the photographs and job titles used to represent each condition does not matter or that the stimuli chosen represent the population of all possible stimuli that one could have used.

The study using the larget and most diverse set of stimuli in the manipulation of sex, attractiveness, and sex-type of job was conducted by Johnson et al (2010). In their study 1, an F2 analysis was conducted in which photograph was used as participant. A set of eight raters evaluated attractiveness of 204 (102 male, 102 female) pictures. Each picture was categorized as attractive or unattractive based on the median split on ratings of attractiveness. Another set of 67 raters evaluated the sex-type of each of 26 jobs. Based on the median rating of sex-type, each job was categorized as a male-type or female-type job. Finally, another set of eight raters evaluated

the suitability of employment of each picture for 4 of the 26 jobs and the mean of ratings of each set of eight raters constituted the dependent measure. A total 96 raters participated in this phase of the research but only the mean ratings of suitability of each picture/applicant was used as the dependent measure. Thus, the picture constituted the subject with ratings aggregated across human participants. Using the picture as subject, a 2 (sex of picture) X 2 (attractiveness of picture) X 2 (sex-type of job) X 2 (importance of appearance to job) within subjects analysis of variance was conducted on the mean ratings of employability suitability for each photograph. The researchers found an attractiveness bias in which attractive candidates were rated more positively than unattractive candidates. and a sex X job type interaction in which males were rated more favorably for male-typed jobs and females for female-typed jobs. The researchers found no evidence of a beauty is beastly effect but did find a three-way applicant sex X attractiveness X job-type interaction in which attractiveness provided somewhat less benefit for women applying to a masculine position. Although Johnson et al (2010) provided a larger sample of stimuli, they did not conduct statistical analyses that incorporated ariability in both stimuli and human participants as recommended by Clark (1973) and Fontenelle, et al (1985).

Despite the frequent admonitions to provide a larger sample of stimuli and to statistically account for stimulus variability, these recommendations are largely ignored in the research on attractiveness biases in the evaluations of job applicants. Pursuant to correcting for the deficiencies in previous research, the present study incorporated three methodological modifications in the test of the hypotheses.

1. A stronger test of the “beauty is good” and “beauty is beastly” hypotheses was implemented by examining the effects of target attractiveness on evaluations for female typed and masculine typed jobs.

2. A larger set of stimuli were used in the manipulations of the primary independent variables than typically used and these stimuli were randomly sampled and assigned to each participant.

3. Statistical tests were used that incorporated both human participant and stimulus variability in the tests of the hypotheses.

Three hypotheses were tested:

Hypothesis 1: Attractive applicants are evaluated more favorably than unattractive applicants.

Hypothesis 2: Consistent with the person-job fit model, applicant sex and and job sex-type are hypothesized to interact in their effects on suitability evaluations. The sex-type of male applicants are evaluated more favorably than female applicants for male sex-typed positions, whereas female applicants are evaluated more favorably than male applicants for female sex-typed positions.

Hypothesis 3: Consistent with the beauty-is-beastly hypothesis, a three-way interaction is hypothesized in which the effects posed in hypothesis 2 are further moderated by applicant attractiveness. Attractive females are evaluated more favorably than unattractive females for strongly female-typed job. Unattractive females are evaluated more favorably than attractive females for strongly male-typed job.

Method

Experimental design

Each participant evaluated the suitability for hire of a job each of ten applicants on the basis of a data sheet. The ten data sheets included the manipulations of attractiveness (moderately high, moderately low), applicant sex (male/female), and gender job (male/female) type and two moderately attractive male control applicants. A 2 (attractiveness) X 2 (applicant sex) X 2 (gender

job type) within subjects design was implemented to test the hypotheses. Each participant evaluated ten applicants for a positions on the basis of a one page data sheet. The manipulation of the independent variables was conducted by means of the following eight data sheets

1. An attractive male appilcant for a male typed job: AMMJ
2. An attractive male applicant for a female typed job:AMFJ
3. An attractive female applicant for a male typed job: AFMJ
4. An attractive female applicant for a female typed job: AFFJ
5. An unattractive male applicant for a male typed job: UMMJ
6. An unattractive male applicant for a female typed job: UMFJ
7. An unattractive female applicant for a male typed job: UFMJ
8. An unattractive female applicant for a female typed job: UFFJ

In addition to these, two other data sheets were provided as fillers to provide a more realistic set of applicants.

9. An moderately attractive male applicant for a male typed job: Filler 1
10 A moderately attractive male applicant for a male typed job: Filler 2

Manipulations of independent variables

Each participant was randomly assigned to each of the eight experimental applicants All participants were given the same two filler applicants. To manipulate applicant sex and attractiveness a photograph was randomly drawn for each condition and each participant from a pool of forty pre-rated photographs. The pool consisted of ten photographs for each of the following conditions: attractive female, unattractive female, attractive male and unattractive male.

To manipulate job gender-type, a job title was randomly selected for each condition and each participant from a pool of twelve pre-rated job titles. The positions were chosen based on pilot research, and to provide a strong test of the job-person fit model, the 12 positions were intentionally selected to represent extremes of sex-typing. Six of these titles had been identified as highly male-typed jobs and the other six as highly female-typed. The six female typed jobs and the perceived percentage of those employed in the position who were female were lingerie salesperson (92.4%), cosmetics salesperson (89.48%), secretary (79.7%), office receptionist (77.72%), director of day care services (77.36%), and executive secretary (76.15%). The six male typed jobs and the perceived percentages of male occupants were car salesperson (79.45%), director of security (80.87%), hardware salesperson (80.95%), prison guard (82.82%), construction supervisor (87.48%), and tow truck driver (89.31%). The male and female positions constituting the pools were approximately the same status levels and each pool consisted of two supervisory positions. In the pilot research the perceived status of the positions were also rated and male and female-typed positions were chosen that were similar. The mean status rating of the six male-typed positions was 3.37 whereas the mean status rating of the six female position was 3.33.

The same photographs, data sheets, and job titles were used to represent the moderately attractive male candidates and the male-typed positions for the two filler applications. A pool of twelve pre-rated names were used in assigning names to applicant data sheets. There were three male male-appropriate names, three female appropriate names, and six gender neutral names. The male photographs were assigned at random a name from the male appropriate and neutral names. The female photographs were assigned at random a name from the female appropriate and neutral names. Evaluations for each applicant on the data sheet were on a five point scale (1

= very poor to 5 = very good) for amount of previous work experience, reference letter from previous employers, personality test scores, interview score, ability test scores, and overall evaluation. All applicants had an overall rating of "5" or very good. Two fives and three fours were assigned to the other five dimensions. The numerical rating was randomly assigned to each dimension on each data sheet. The order in which the ten data sheets were presented to each participant was randomly determined.

Procedures

The participants were given the following instructions:

> "On each of the following pages you see a photograph of a person who has applied for a job. The photograph and the title of the job for which the person has applied is presented along with a seven-point scale on which you are to rate the employment suitability of the person photographed for the job."
>
> "The candidates that you will see passed an initial screen. You will see the evaluation of the HR manager of the amount of work experience, personality test scores, ability test scores, interview score, employer references, and overall evaluation for each candidate. Each applicant evaluated was described as having been prescreened in previous testing and interviewing sessions. One of the scale is a very low score while five represents the best score. Click the point on the scale that reflects how you would rate the person on the scales provided."

Participants evaluated each applicant on how suitable for employment the applicant was for the position on a seven point scale that ranged from 1 (not at all) to 7 (extremely). The order in which the eight applicants representing the experimental conditions and the two filler applicants was randomly determined. There was no time limit placed on the participants in their

evaluation of each applicant. Once they had provided a rating, they were to click a button to continue to the next applicant but once evaluated, the participant could not go back and change their rating of an applicant. After they had finished evaluating all ten applicants, the participants were asked to pick the one applicant who provided the best fit to the position and the one applicant that represented the worst fit. They were allowed to move back and forth among the applicants in making these best-fit and worst-fit choices. Participants were asked to indicate their sex, whether they had experience in hiring applicants (yes or no), their ethnicity (white vs nonwhite), and their age (in years). A preliminary analysis indicated that no main effects or interactions with the primary independent variables were found for these demographic characteristics. Consequently, the effects of these variables are not reported in the tests of the hypotheses.

Participants

A total of 375 individuals participated in the research. Of these, …… were recruited via Mturk and compensated for their participation. The remaining participated as part of a classroom exercise. Of these participants, 138 indicated that they had experience hiring whereas 237 indicated that they had not. Twohundredand eightysix of the participants were white, 43 were black, 26 were Asian and 20 were of some other ethnicity. The mean age of the participants was 29.95 years and ranged from 18 to 68.

Results

Analyses.

Clark (1973) and others (Fontenelle et al, 1985; Judd, et al, 2012) argue that a fixed effect ANOVa in which only human subjects are treated as a random variable is biased and does not tell us whether the effect generalizes across the stimuli sampled. In some cases statistically significant effects could be the result of a few of the photographs or job titles that were used in

the manipulation of the conditions. In addition to the F1 analysis which is typically used in testing the effects of attractiveness, sex, and gender-typing, one needs to compute F2 in which stimuli are treated as subjects and evaluations of applicants are averaged across participants given the same combination of photograph and job title. Assuming that there is a sufficient number of stimuli to allow for a powerful test, F2 should allow one to assess whether effects generalize across stimuli. A third analysis is the computation of the minF' which allows one to assess whether effects generalize across both stimuli and participants (MinF'= F1F2/(F1 + F2)). The use of minF' is not without controversy with some claiming that it is unduly conservative (see Raaijmakers, Schrijnemakers & Gremmen, 1999 and Fontenelle et al, 1985 for discussions of the issue). However, it remains the most commonly used means of assessing generalizability of effects across both participants and stimuli.

Tests of hypotheses

To test the hypotheses , our sets of analyses were conducted in testing the main effects and interactions of attractiveness, sex, and job-sex type. The results of these analyses are reported in table 1. The F1 tests consisted of a three way within subjects analysis of variance treating attractiveness, sex, and job-type as repeated measures factors and using human participants as subjects. Three other analyses were conducted to compute F2s. In the first of these, a three way ANOVa was conducted using the individual pictures as subjects. In this analysis attractiveness and sex of applicant were between-group factors and job sex-type was a repeated measures factor. Another three-way ANOVa was conducted using the twelve job titles as subjects. In this analysis attractiveness and sex of applicant were repeated measures factors and job sex-type was a between-subject factor. Finally, a three-way ANOVa was conducted using the picture-job title combinations as subjects. Here applicant sex, applicant attractiveness, and

job sex-type were all between subject factors. For each of the F2 analyses, a quasi-F was computed that took in account variability across both human participants and stimuli. In addition to these primary tests of the hypotheses, Chi square tests were conducted on the distribution of best and worst choices to assess the effects of attractiveness, sex, and job sex-type.

Table 1

Statistical Tests of Hypotheses Using People, Pictures, Jobs,

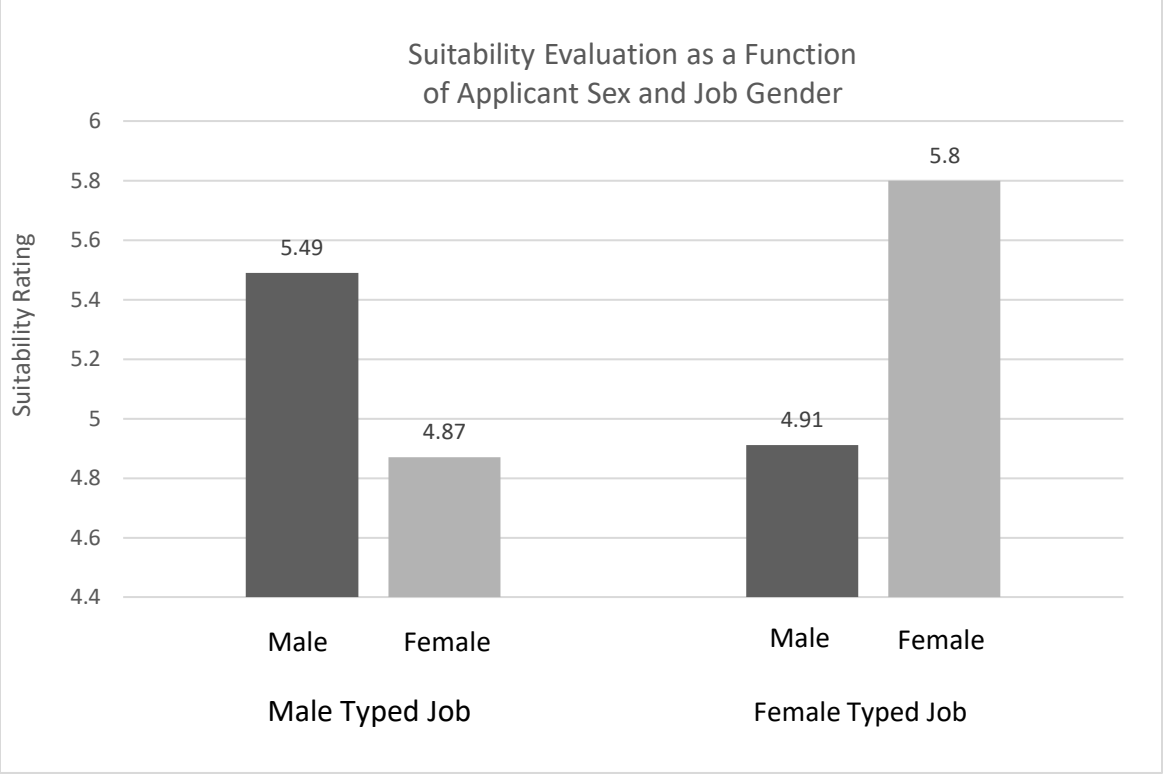
and Picture-Job Combinations as Ss

	People as Ss	Pictures as Ss		Jobs as Ss		Picture-Job Combinations as Ss	
Source	F1	F2	minF'	F2	minF'	F2	minF'
Attract.	25.49**	19.05**	10.91**	19.43**	11.02**	10.25**	7.31**
Sex	13.48**	8.01**	5.02*	.88	ns	4.94*	3.62^{+}
Job Gender	25.33**	8.92**	6.60*	.98	ns	11.46**	7.89**
Attract. X Sex	.05	.05	ns	.04	ns	.16	ns
Attract. X Job	1.33	.04	ns	1.19	ns	2.04	ns
App. Sex X Job	199.84**	168.32**	91.37**	29.31**	23.05**	136.70**	81.17**
Attract. X Sex XJob	2.18	.99	ns	1.54	ns	.34	ns

F1 analyses . Four statistically significant effects were found in the 2 X 2 X 2 repeated measures analyses used to compute the F1s. A main effect was found for attractiveness in which attractive applicants were evaluated as more suitable for the position (M = 5.37) than unattractive applicants (M = 5.17), $F(1, 375) = 27.574$, $p < .001$, $\eta^2 = .065$). An applicant sex X job sex-type interaction was also found and is depicted in the figure below ($F(1,375) = 199.837$, $p < .001$, , η^2

= .348) . When evaluating the applicants for a male typed job, the male applicant was evaluated as more suitable for the position than a female applicant (M = 5.485 vs. 4.872). When evaluating the applicants for a female-typed job, the female applicant was evaluated as more suitable for the position than a male applicant (M = 5.803 vs 4.902).

Figure 1



F2 and Quasi-F Analyses. To assess the generalizability of effects across stimuli another set of analyses were conducted using the recommendations by Clark (1973). First, individual picture were used as subjects and an ANOVA was conducted in which the attractiveness and gender factors were treated as between group factors and job-type as a repeated factor. The same statistically significant effects found in these F2 analyses as found in the F1 analyses. There were main effects for attractiveness (F2 (1,36) = 19.05, $p < .001$, $\eta^2 = .346$), applicant sex (F2 = 8.01, 1/36, $p < .05$, $\eta^2 = .182$), and job sex-type (F2 (1,36) = 8.92, $p < .001$, $\eta^2 = .199$), as well as a

applicant sex X job sextype interaction (F2 = (1,36) = 168.32, $p < .001$, $\eta^2 = .892$). The minF' was statistically significant for attractiveness (minF' (1,104) = 10.91, $p < .01$), applicant sex (minF' (1, 88) = 5.02, $p < .05$), sex type of job (minF' (1,65) = 6.60, $p < .05$), and the applicant sex X job sex type interaction (minF' (1,114) = 91.37, $p < .01$).

In another analysis job titles were used as subjects and an ANOVA was conducted in which attractiveness and gender were repeated measures factors and job-type was a between-groups factor. A statistically significant main effect for attractiveness was found (F2(1,10) = 19.43, $p < .001$, $\eta^2 = .66$) as well as a statistically significant interaction between job type and applicant sex (F2(1,10) = 29.31, $p < .01$, $\eta^2 = .746$). The main effect for sex type of job found in the F1 analysis was statistically nonsignificant (F2(1,10) = .981, $p < .35$). The quasi-F was statistically significant for attractiveness (minF' (1,31) = 11.03, $p < .01$) and the applicant sex X job type interaction (minF' (1,15) = 23.55, $p < .01$).

In the third F2 analyses, the job-title + picture combinations were treated as participants and the ANOVa treated the attractiveness, sex, and job sex-type as between subject factors. Statistically significant effects were found for attractiveness effect (F2 (1,471) = 10.253, $p < .001$, $\eta^2 = .021$), applicant sex (F2 (1,471) = 4.936, $p < .05$, $\eta^2 = .01$), sex-type of job (F2 (1,471) = 11.458, $p < .05$, $\eta^2 = .024$) and the interaction of job sex type and applicant sex (F2 (1, 471) = 136.70,, $p < .001$, $\eta^2 = .225$. The Quasi-Fs that were calculated were statistically significant for the attractiveness main effect (minF' (1,770) = 7.31, $p < .01$), job (minF' (1, 790) = 7.89, $p < .01$), and job X sex (minF' (1,841) = 81.17, $p < .01$). The effect for applicant sex was marginally significant (minF' (1,753) = 3.62, $p < .06$),

Chi square analyses of best and worst fit choices. Table 2 summarizes the frequencies of choices as best and worst fit.

Table 2

Distributions of Applicants Picked as

Best and Worst Fits to the Position

Applicant and Job Type	Best Fit		Worst Fit	
	Frequency	Percent	Frequency	Percent
Attractive Female/Female Job	128	34.0	12	3.2
Attractive Female/Male Job	19	5.1	54	14.4
Attractive Male/Female Job	27	7.2	56	14.9
Attractive Male/Male Job	58	15.4	18	4.8
Unattractive Female/Female Job	40	10.6	29	7.7
Unattractive Female/Male Job	13	3.5	46	12.2
Unattractive Male/Female Job	13	3.5	62	16.5
Unattractive Male/Male Job	34	9.0	17	4.5
Controls	44	11.7	82	21.8

The chi square testing the differences between the observed distribution of best fit choices and the expected distribution of each of the experimental applicants (i.e., .10 for each of the eight applicants) and the two controls (i.e., .20 for the two controls) was statistically significant (Chi square = 286.23, df = 8, $p < .001$). The chi square for the difference between the distribution of observed worst fit choices across the eight experimental conditions and the controls and the expected distribution of worst fit choices (i.e. .10 for each of the eight applicants and .20 for the controls) also was significant (Chi square = 75.38, df = 8, $p < .001$).

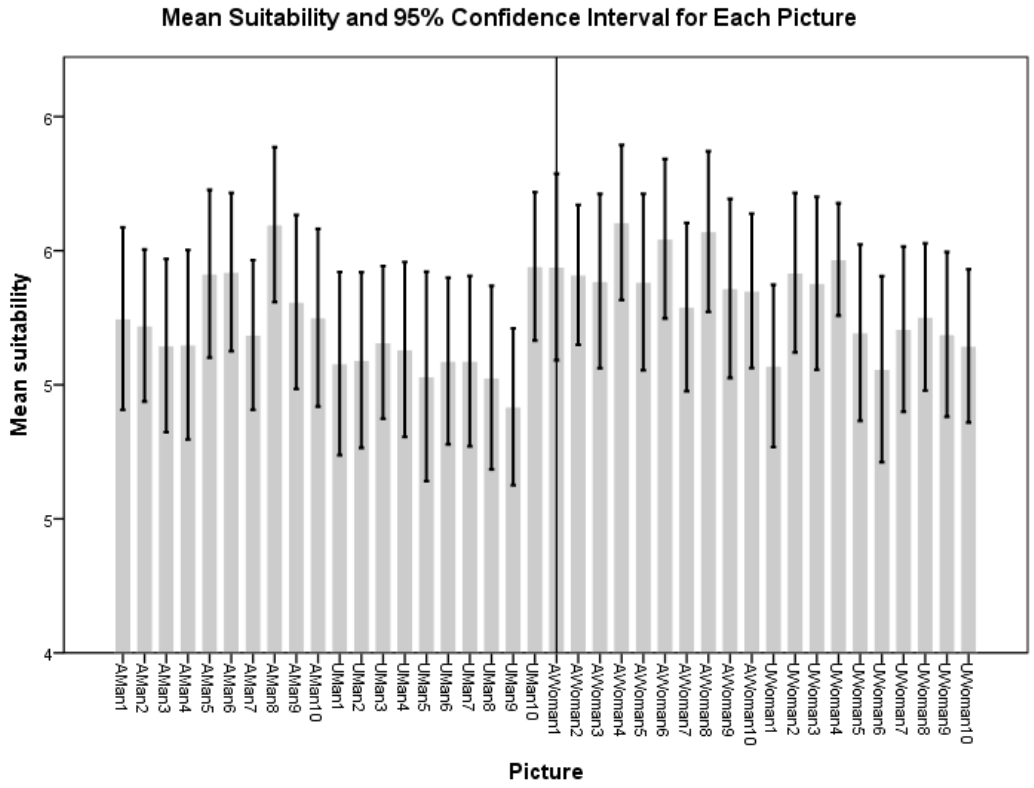
The attractive candidate was chosen as the best fit by 232 of the 375 participants (62%), whereas 100 chose the unattractive candidate as the best fit (27%) and 43 chose the control or moderately attractive male candidate (11%). The chi square testing the differences between the observed distribution of best fit choices and the expected distribution of attractive, unattractive, and control applicants (i.e., .40, .40, and .20, respectively), was statistically significant (Chi square = 74.11, df = 2, $p < .001$). The chi square testing the difference between the observed distribution of worst fit choices and the expected distribution of attractive, unattractive and control applicants (i.e., .40, .40 and .20, respectively) did not achieve contentional levels of significance (Chi-square = 1.42, df =2, $p < .4917$).

A woman was chosen as the best fit by 200 participants (53.2%) and 176 chose one of the six male applicants (47.81%). The difference in the observed and expected distributions of best fit choices was statistically significant (Chi-square = 27.26, df =1, $p < .001$). A female applicant was chosen as the worst fit by 141 participants (37.5%), whereas 235 chose a male applicant as the worst fit (62.67%). The difference between the observed and expected distributions of worst fit choices was statistically nonsignificant (Chi square = .979, df = 1, $p <.3225$).

Exploratory analyses

The results of the F1, F2, and quasi-F analyses provided no support for a beauty is beastly effect and indicated that the variability across stimuli swamped any systematic bias in favor of unattractive females for male jobs. This led to the question of whether the beauty is beastly effect was more likely to occur for some specific pictures and jobs or whether the effect was so rare as to be essentially absent in this dataset. As seen in figure 2, there was considerable variability in the suitability ratings given to each picture used in the manipulations of attractiveness and gender and it is possible that for some specific pictures a beauty is beastly effect would emerge.

Figure 2



AMan 1 – 10: Attractive Male
UMan 1 – 10: Unattractive male
AWoman 1 – 10: Attractive Woman
UWoman 1 – 10: Unattractive Woman

A first set of analyses focused on just those participants who gave lower suitability ratings to an attractive woman applying for a male-typed job than to an unattractive women applying for a male-typed job. This beauty is beastly effect was found in 114 cases. The most frequently represented job titles for the attractive female/male-typed job condition in these 114 cases were prison guard (32), tow truck driver (26) and hardware salesperson (21). The least represented were car salesperson (7), construction supervisor (14) and director of security (14). A chi square test of the discrepancy between this distribution and an expected distribution in which all six titles were equally represented was statistically significant (Chi square = 21.90, df =

5, $p < .001$). Additionally, the F1 ANOVAs were repeated separately for the participants given each of the six male sex-typed positions in the attractive female/male typed job condition (see table 3). Evidence of a beauty is beastly effect was found in the form of a three-way interaction for the positions of prison guard and tow truck driver. In each case, the unattractive female candidate applying for the male-typed job (i.e., prision guard or tow truck driver) was evaluated as more suitable than the attractive female candidate applying for one of the other male-typed jobs. A three way interaction was also found for car salesperson and director of security, but in these two cases the attractive female was evaluated as more suitable than the unattractive female.

Table 3

Job Titles Associated with Beauty is Beastly Effects

Job in Attractive Female/Male Job Condition	Attract.XSexXJobtype Interaction Effect	Attractive Woman/ Male Job	Unattract. Woman/ Male Job	t-test (one-tailed)
Car Salesperson	$F(1, 63) = .49$, ns	5.67	4.98	$t(63) = 3.83$, $p < .01$
Construction Supervisor	$F(1, 52) = .19$, ns	4.94	4.81	$t(52) = .79$, ns
Director of Security	$F(1, 60) = .47$, ns	5.03	4.48	$t(60) = 2.34$, $p < .01$
Hardware Salesperson	$F(1, 71) = .03$, ns	5.04	4.85	$t(71) = 1.34$, ns
Prison Guard	$F(1,62) = 16.43$, $p < .01$	4.27	4.95	$t(62) = 4.84$, $p < .01$
Tow Truck Driver	$F(1, 61) = 8.53$, $p < .01$	4.55	4.85	$t(61) = 1.91$, $p < .05$

Similar exploratory analyses were conducted for the 40 pictures to determine the conditions in which a beauty is beastly effect was most likely to occur. Again, focusing only on those 114 cases in which an unattractive female was rated higher than an attractive female for a male-type position, the frequency with which each of the 10 female attractive pictures used in this condition was calculated. A chi square test of the discrepancy between this distribution and

an expected distribution in which all ten were equally represented was statistically nonsignificant (Chi square = 9.33, df = 9, p < 408). In another exploratory anaylsis the frequency with which each of the 10 unattractive female pictures was present in the 114 beauty-is-beastly cases was examined. A chi square test of the discrepancy between this distribution and an expected distribution in which all ten of the unattractive female pictures was equally represented was statistically nonsignificant (Chi square = 8.11, df = 9, p < .53). As was done for the job titles, the F1 analyses were repeated separately for the participants given each of the ten attractive female pictures in the attractive female/male typed job condition. Of the twenty ANOVAs, only two revealed significant 3-way interactions and in one case there was a beauty is good effect and in the other a nonsignificant beauty is beastly effect. In general the beauty is beatly evidence does not appear to occur as a function of the individual pictures used in this experiment.

Discussion

In support of the hypotheses, a bias against unattractive candidates was found in the evaluations of their suitability for positions. This bias was found regardless of the sex-typing of the position or the sex of the applicant. The strongest effect was an interaction of applicant sex and position sex-typing in which male applicants were rated as more suited to male-typed positions and female applicants were rated as more suited for female-typed positions. This effect was not surprising given that the titles used in the manipulation of job sex-typing represented stereotypic male and female positions. The use of strongly sex-typed positions was a pre-requisite to providing a strong test of the person-job fit interpretation of the beauty is beastly effect. The person-job fit interpretation posits that evaluators attribute more feminine traits to a highly attractive female and more masculine traits to a highly unattractive female. These attributions lessen the perceived fit of the attractive female and enhance the perceived fit of the

unattractive female physical attractiveness when the position is male-typed and presumably requires stereotypic masculine traits. A problem in previous research on the beauty is beastly effect was the use of positions that were not clearly sex-typed. In providing an unequivocal manipulation of sex-typing, the present study casts some doubt on the robustness of the beauty is beastly effect. The analysis of the best and worst choices suggested that the bias in favor of attractive candidates was most pronounced in the best choice picks and the candidate most preferred were the females applying for female-typed positions. Consistent with the conclusions of previous meta-analyses, physical attractiveness appears to provide a consistent advantage for both males and females applying to both male-typed and female-typed positions (Hosada et al, 2003).

This was the first empirical test of attractiveness effects to apply the suggestions of Clark (1973), Fontenelle, Phillips & Lane (1985), and others for incorporating stimulus variability into tests of hypotheses. The attractiveness bias appeared robust and generalized across human participants, position, and picture used in the manipulation of applicant sex and attractiveness. The effect was not especially strong given that all applicants were depicted as highly qualified for the positions for which they were applying. However, the bias in favor of the attractive applicants was consistently found in the F1, F2, and quasi-F analyses. There was no evidence that the sex of the rater or applicant moderated these effects. A bias against unattractive applicants was found regardless of whether the rater and ratee were male or female.

Although the beauty is good effect was strongly supported, these findings do not entirely negate previous research showing a beauty is beastly effect. The evidence seems clear that attractiveness can constitute a disadvantage for female applicants in specific circumstances. That the effect “can” occur for specific stimuli suggests that the likelihood with which the effect

occurs is perhaps lower than implied by previous investigators who have reported beauty is beastly effects. In the exploratory analyses we found that the bias in favor of attractive candidates was consistent across pictures but that the effect was most likely occur in evaluating women for some male-typed jobs. The two positions in which a beauty is beastly effect was most likely to occur was for tow-truck driver and prison guard. It should be noted that although an attractive woman working in these jobs remains a novelty, it does occur. Indeed, Title VII of the Civil Rights Act prohibits discriminating against the hiring of correctional officers on the basis of sex, and the proportion of women working as corrections officers has dramatically increased over the last few decades. The Bureau of Justice Statistics study Women in Law Enforcement, 1987-2008, published in June 2010, shows that in 2008 14 percent of Bureau of Prisons officers were female.

Not only can a beauty is beastly effect occur, but we would further suggest that the use of single or limited stimulus samples is appropriate if researchers are attempting test theoretical explanations for an effect. The most important consideration in a test of theory is to evoke the effect and then manipulate the conditions under which it is more or less likely to occur. Consequently, it is appropriate to set up a situation in which stimuli are used that are likely to a bias against attractive applicants so as to flush out the phenomenon and identify the antecedents, moderators, and mediators of the effect. Akin to an entymologist in search of a rare species of butterfly would need to capture this rare type to understand it, a researcher need to create a situation in which a beauty is beastly effect is likely to occur to fully understand the antecedents, mediators, and moderators of the effect. It is not appropriate, however, to draw conclusions for the prevalence of an effect when a narrow set of stimuli are used and no attempt is made to assess the effects of stimulus variability. An entymologist might be able to capture a rare

butterfly so as to study it, but the successful capture does warrant the conclusion that it is no longer rare. Similarly, the use of single or limited stimulus samples is appropriate to capture "effects" and to see if they "can" occur and the circumstanes under which they occur but are inappropriate to assess the generalizability of the effects across a more realistic and varied set of stimuli. Stimulus sampling has important implications for a wide variety of lab and field, correlational and experimental research and needs to receive more attention from researchers. Stimulus sampling is a concern dating at least as far back as Ego Brunswik and is relevant to any research using stimuli to manipulate independent variables. For example, it is not uncommon to create scenarios to manipulate variables important to testing hypotheses derived from social justice or equity theory. This author has never seen an instance in which attempts were made to provide a diverse sample of scenarios in the manipulation of variables and to incorporate this variability among scenarios into statistial analyses.

The present study demonstrated that the beauty is beastly effect can occur but does not appear especially robust in its generalization across stimulus conditions. Nonetheless research is need to explore conditions in which it is most likely to occur. One such potential condition is where the attractive other is seen as a rival. There is some research showing that when there is an expectation of future interaction, female raters are more likely to show a beauty is beastly effect in their evaluations of a same-sex applicant (Luxen & van de Vijver, 2006). Other research suggests that the beauty is beastly effect is mitigated when the attractive female applicant acknowledges that her appearance may not seem to fit the position (Johnson, Sitzmann & Nguyen, 2014) and when the rater is high in attractiveness and the threat of the other's attractiveness is presumably low (AgtheSpörrle & Maner, 2010). It should be noted, however, that these and other researchers reporting a beauty is beastly effect used small samples of stimuli

in their manipulations and did not conduct statistical analyses incorporating variability across both human participants and stimuli. The present study corrected for both of these omissions and based on the results it seems warranted to conclude that a bias in favor of attractive applicants is quite robust and that the beauty is beastly effect is limited to a small domain of male-typed jobs.

References

Agthe, Maria; Strobel, Maria; Spörrle, Matthias; Pfundmair, Michaela; Maner, Jon K. (2016). On the borders of harmful and helpful beauty biases: The biasing effects of physical attractiveness depend on sex and ethnicity. *Evolutionary Psychology*,14, ???

Agthe, M., Sporrle, M., & Maner, J. K. (2011). Does being attractive always help? Positive and negative effects of attractiveness on social decision-making. Personality and Social Psychology Bulletin, 37, 1042–1054.

Agthe, M., Spörrle, M., & Maner, J. K. (2010). Don't hate me because i'm beautiful: Anti-attractiveness bias in organizational evaluation and decision making. *Journal of Experimental Social Psychology, 46,* 1151-1154. *doi:10.1016/j.jesp.2010.05.007*

Banducci, S. A., Karp, J. A., Thrasher, M., & Rallings, C. (2008). Ballot photographs as cues in low-information elections. *Political Psychology, 29,* 903-917. *doi:10.1111/j.1467-9221.2008.00672.x*

Bzdok, D. D., Langner, R. R., Caspers, S. S., Kurth, F. F., Habel, U. U., Zilles, K. K., & ... Eickhoff, S. B. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure & Function, 215,* 209-223. *doi:10.1007/s00429-010-0287-4*

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal Of Verbal Learning & Verbal Behavior, 12*, 335-359. doi:10.1016/S0022-5371(73)80014-3

Bzdok, D. D., Langner, R. R., Caspers, S. S., Kurth, F. F., Habel, U. U., Zilles, K. K., & ... Eickhoff, S. B. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure & Function, 215, 209-223. doi:10.1007/s00429-010-0287-4*

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109-128. *doi:10.1037/0033-2909.110.1.109*

Eastwick, P. W., Luchies, L. B., Finkel, E. J., & Hunt, L. L. (2013). The predictive validity of ideal partner preferences: A review and meta-analysis. *Psychological Bulletin, ??? doi:10.1037/a0032432*

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304-341. doi:10.1037/0033-2909.111.2.304

Fontenelle, G. A., Phillips, A. P., & Lane, D. M. (1985). Generalizing across stimuli as well as subjects: A neglected aspect of external validity. *Journal of Applied Psychology, 70*, 101-107. doi:10.1037/0021-9010.70.1.101

Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods, 12*, 554-566. doi:10.1177/1094428107300396

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). *Personnel Psychology, 56(2),*. The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. 431-462. doi:10.1111/j.1744-6570.2003.tb00157.xA

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly, 58, 108-122. doi:10.2307/2787149*

Johnson, Stefanie K.; Sitzmann, Traci; Nguyen, Anh Thuy. (2014). Don't hate me because I'm beautiful: Acknowledging appearance mitigates the "beauty is beastly" effect. *Organizational Behavior & Human Decision Processes, 125*, 184-192.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*, 54-69. doi:10.1037/a0028347

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126(3), 390-423. doi:10.1037/0033-2909.126.3.390*

Luxen, M. F., & van de Vijver, F. J. R. (2006). Facial attractiveness, sexual selection, and personnel selection: When evolved preferences matter. *Journal of Organizational Behavior, 27,* 241–255.

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24,* 1315-1344. *doi:10.1111/j.1559-1816.1994.tb01552.x*

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416-426.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review Of Psychology, 57,* 199-226. *doi:10.1146/annurev.psych.57.102904.190208*

Ritts, V., Patterson, M. L., & Tubbs, M. E. (1992). Expectations, impressions, and judgments of physically attractive students: A review. *Review Journal of Educational Research, 62*, 413 - 426.