Effects of Beauty Filters on the Perception of Job Qualification

Spring 2022 - W241 Final Report

Meer Wu, Joyce Li, Amy Jung

Abstract

Do beauty face filters affect the perception of job qualification? We are curious to investigate whether artificial beauty generated from face filters can affect the perception of job qualification, which we define as having the required experience, skills, or attributes a position requires. In this conjoint experiment, respondents are asked to select between two randomly-generated candidates, each applying to a different role, having different backgrounds, and being assigned a random filtered or unfiltered photo. Our results indicate that beauty face filters do not have a statistically significant effect on being chosen for a job, while being white and having more years of experience significantly impact one's chance at landing the job. Through this experiment, we conclude that face filters do not significantly improve perception of one's job qualifications.

Introduction

Our team is investigating whether attractiveness is a significant factor in evaluating job qualification, and whether augmentation of our attractiveness through changes in our appearance also has an effect on one's perceived qualifications. We hypothesize that our treatment of FaceTune "beauty filters" will work because it directly allows us to assess whether conventional beauty standards of attractiveness affect perception of job qualification. We reason that attractiveness is the mechanism that may boost human perception of competence. Therefore, by changing the facial shape and features (ie. brighter skin, less blemishes, smaller nose, sharper chin, etc.) to augment the profile's beauty, we would be controlling the attractiveness factor that feeds into the perception of job qualifications.

Background

Beauty filters, automated photo editing tools that use artificial intelligence and computer vision to detect and change facial features, have been increasing in popularity in recent years. Editing apps such as FaceTune, which experienced usage increase by 20% the past few years and has over one million edited images exported daily, allow users to smooth, shrink, and sharpen their features at the tip of their fingers [1]. According to news articles, beauty filters may be changing the way we see ourselves [2].

In regards to the job market, past research has indicated that women who wear makeup and jewelry in society are seen as more competent than women without them [3]. In another study, attractive people were found to have a higher chance of being contacted by their employers [4]. However, other studies investigating the effect of attractiveness and qualification on a job applicant's hiring decision concluded that qualifications, rather than attractiveness, had a greater influence on a manager's perception of the candidate's suitability for the position [5]. Considering the lack of consensus on this subject and no studies performed on the effects of beauty filters on job qualification, we further investigate the effect of attractiveness on perceived qualifications with the beauty filter app, FaceTune 2.

Research Question

Our research question is: Do beauty face filters affect the perception of job qualification?

We hypothesize that face filters will improve the perception of one's job qualifications, which could be beneficial when seeking job opportunities. More specifically to our experiment, we hypothesize that the effects would be more prominent in women than in men, and also in front-facing roles than in back-facing roles. Therefore, the outcome we predict is that the treatment of beauty filters will increase the probability of being chosen when two profiles of job applicants are shown.

Randomization Process

Using Qualtrics, we were able to create a conjoint experiment and complete randomness for all candidate features. This also means that each survey respondent could be in both the control and treatment group for this experiment, depending on whether they were randomly presented the filtered images or not. Following this paper, we wrote a JavaScript Fisher-Yates shuffle function at the top of our survey that would randomly choose one level of each feature to display for each candidate when the question loaded [7]. We then created an HTML table to display both candidate choices (and all randomized features) to the respondent. For each question, the table refreshed with a new set of randomized features generated from the JavaScript function. For photos specifically, we randomized the selection between 24 photos because there was no way of tagging photos with gender, ethnicity, and filters in Qualtrics. We made sure that the 24 photos were representative of all possible combinations of filter, gender, and ethnicity (3 photos of each gender x ethnicity pairing = 2 x 2 x 3 = 12 x 2 (filtered photos for each gender/ethnicity pairing) = 24 total photos.) To ensure that this randomization was implemented correctly, we made sure to manually try looking at multiple sets of randomized candidates to ensure that attributes were randomized.

We used Amazon Mechanical Turk in order to get responses for our survey. Because survey participants are not randomly chosen, we wanted to perform a covariate balance check with our data after collection to ensure that our survey wasn't only being taken by people with the same characteristics.Using a regression model with all randomized features and respondent demographics, we conducted a covariance balance check. None of the covariates had significant effects on the outcome, and an F-test comparing a model with covariates and the model without them confirmed that the covariates indeed had no impact on the results.

Treatment

Our treatment of interest is FaceTune 2 filtering. To ensure that the treatment effect was consistent across all photos, we filtered each photo to the same effect magnitude. Each effect was measured on a scale of 1-100 on FaceTune 2, and we used the following effects: whitened teeth (50), smoothed entire face (50), concealed eyebags (50), increased vibrance on the entire face (50), increased smile (50), enlarged eyes (25), slimmed down nose (50), and thinned jawline (50). See appendix A for examples of unfiltered versus filtered photos. Respondents will be able to see 5 questions with a different randomized image, college, job title, and years of experience. Any photo being shown without a filter is considered a control, and any photo being shown with a filter is considered treatment.

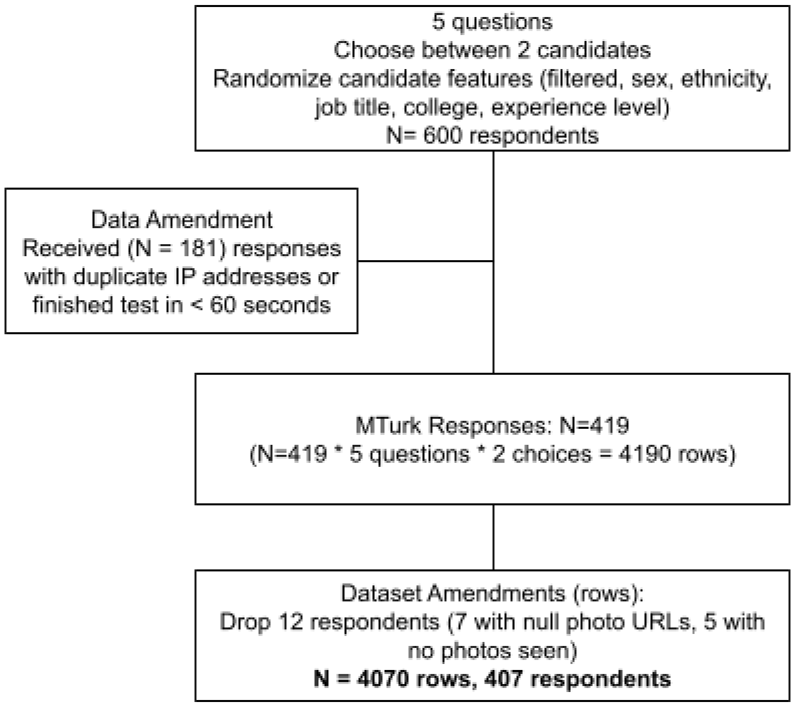
It was uncertain whether all images loaded for all questions per respondent. To ensure that we were measuring the correct treatment effect, we decided to drop the photo choices of 7 respondents who did not seem to have a valid photo url shown to them. We also dropped respondents who took under 1 minute on the survey and those with duplicate IP addresses.

Data Completeness

As shown in the flow diagram below, we first collected 600 respondents on MTurk with the expectation that not all would give valid answers. MTurk provides the functionality to approve or reject respondents. Through analyzing our responses, we found 181 respondents who either had duplicate IP addresses, implying that the same respondent took the test twice, or spent less than one minute on their survey, implying that they randomly clicked through the survey without thought. We dropped these respondents from our dataset, leaving 419 respondents. After this step, we additionally found 7 respondents who had null photo URLs and 5 respondents who reported seeing no photos show up in their surveys. Because these respondents were not able to receive treatment, we dropped them from the dataset as well. We had no risk of attrition since the survey forced respondents to give an answer to all questions. Ultimately, we ended up with 407 respondents and 4070 data points (5 choices x 2 questions per respondent) to analyze.

In order to confirm that respondents did indeed receive treatment, we included a question in the survey that allowed respondents to indicate whether or not they were able to see photos in each question. We expected respondents to either be able to see all photos in the survey or none at all, due to each respondent's personal cookie settings. However, in analyzing responses, we noticed that a majority of respondents indicated that they saw a subset of photos. These answers seemed highly likely to be due to random clicking, so to test this out, we ran two regressions, one where we kept only the questions where the respondent explicitly mentioned that photos were seen, and one where we kept all questions when the respondent reported seeing at least one photo. The model with all questions reported a similar coefficient as the other model but with smaller standard errors, confirming our suspicion that a respondent has seen all photos when they reported seeing at least one. We then proceeded our analysis with the assumption that a respondent who has seen at least one photo has seen all photos in all 5 questions.

CONSORT document



Power Calculation

Given that our effect size is measured in percentages, we can assume that the effect size will be at least 1%. Ideally, the standard deviation to the outcome variable should be at least two times the effect size. Using this power calculator tool and a max sample size of 2000, we were able to calculate that we would need 208 respondents to get at least 95% statistical power. To double check our assumptions, we checked our statistical power with another source, which states that the sample size n must be greater than $\frac{1000 * \ \# \, max \ levels}{\# \, questions * \# \, alternatives \ per \ question}$. For this experiment, this value is $\frac{1000 * 3}{5 * 2}$ = 300 respondents. According to both statistical power tests, our final number of 407 respondents is adequate to have sufficient predictive power in the experiment [6].

Regressions / Results

We conducted three regressions to observe the effect of beauty face filters on the probability of a candidate being chosen. First, we regressed the outcome on just the face filter indicator to observe a vanilla relationship between the treatment and the outcome. Results of this regression are displayed as Model 1 in Table 1. We reported cluster standard errors to account for individual biases in the perception of job qualification. Without taking anything else into account, having a beauty face filter increases a candidate's chance of being chosen by 1%, but this effect is non-significant. Model 2 in the Table 1 shows a regression of the outcome on all randomized features — filtering, ethnicity, sex, the type of role the candidate is applying to, the college they graduated from, and years of experience. Including all randomized features essentially allows us to use the ATE as the Average Marginal Component Effect (AMCE), which measures the effect of each feature individually on the probability of a candidate being selected, averaging across the effects from all other features. Assuming all other randomized features are constant, face filtering still has the same non-significant effect on the outcome as in Model 1. All other randomized features, except the college variable, had significant effects on the outcome. Assuming all else equal to the baseline group, male candidates have a 41% chance of being chosen, while female candidates have a 45% chance of being chosen. Interestingly, we observed that ethnicity and years of experience both played a huge role in one's probability of being chosen. A white candidate, as opposed to a non-white candidate, has a 7% higher chance of being chosen, and candidates with 7 years of experience are 10% more likely to be chosen than candidates with 5 years of experience. This is consistent with how we know the hiring process should and should not be. More experience should make one more qualified for a job, but past research has also shown that White candidates, as opposed to Black candidates, are more likely to be hired [8].

Lastly, we added two interaction terms to Model 2 to create Model 3 since we hypothesized that the effect of filtering will be larger in women than in men and in front-facing roles rather than back-facing roles. After adding the interaction terms, the baseline group, representing non-White females with 5 years of experience who graduated from a top 50-100 ranking college applying to back-facing roles using an unfiltered photo, has a 43.4% chance of being chosen. The effect of filtering is increased to 4.6%, but it is still not statistically significant based on a 0.5 p-value cutoff. The increase in the filter coefficient after taking into account interaction between sex and filtering suggests that beauty face filters do affect men and women differently. More specifically, compared to the baseline group, women benefit from having their photos filtered while men's chances of being chosen are lowered when they filter their photos. However, these results are non-significant, so we further investigated this hypothesis by running two regressions, one on the subset of data with only male candidates and the other on only female candidates, as shown in Table 2. As expected, female candidates are % more likely to be chosen with a filtered photo (*see Model 1 in Table 2*). Conversely, a filtered headshot actually harms a male candidate's chance of being chosen as the more qualified candidate (*see Model 2 in Table 2*). These results are non-significant, as the ATEs are not definitively non-zero and confidence intervals of the two groups overlap. Contrary to our hypothesis, the effect of filtering is essentially the same among front-facing roles and back-facing roles, and both effects are non-significant. The college the candidates graduated from had consistently non-significant effects on one's likelihood of being chosen across all models. This could be a result of either this variable does not affect one's chance of being chosen or the fact that all of our candidates had 5 to 7 years of experience, and at that point in one's career, which college they graduated from did not really matter.

Based on the results of all of the models, we fail to reject the null hypothesis that beauty face filters have no impact on one's chance of being chosen as more qualified for their job. This is against our initial hypothesis that the filters will increase one's chance of being chosen, but the significant effects of ethnicity and experience are consistent with our understanding of hiring standards and biases during the hiring process.

Limitations & Future Enhancements

There are several limitations of this study that may have created some bias or decreased the generalizability of the study's results. First, the survey takers were all from Amazon Mechanical Turk (MTurk). From our demographic questionnaires at the end of the conjoint analysis, we found that this population (N = 407) comprised of individuals who identified as 88% (n = 357) White, 62% (n = 253) had no formal education past high school, 62% (n = 251) male, and 44% were between the ages of 25 and 34. This specific population may not be wholly representative of real-world recruiters, who typically review the candidates' job resumes. Another limitation with MTurk is that survey respondents may have randomly clicked to answer questions throughout the survey, as mentioned in the Data Completeness section.

In addition, the wording on the surveys may have confused survey respondents, with "For each pair of candidates, select the candidate you think is more qualified for their respective position" either interpreted as whether the candidate was qualified for the job they already had or whether the candidate was qualified for the job they were applying for. Despite the survey starting with the description: "There are 5 pairs of candidates *applying* for various job positions. For each pair of candidates, select the candidate you think is more qualified for their respective position.", survey respondents may have skipped the introduction. Also, the candidate profiles were not full resumes, and did not list specific job experiences or college majors, which may hinder generalizability of this study to the real job market.

In the future, we would recommend creating resumes with LinkedIn profiles, submitting them to actual job listings and reaching out to recruiters. By measuring engagement in the form of interview requests and replies from recruiters as the outcome to determine if beauty face filters affect perception of job qualification, future studies could improve upon the limitations of this experiment.

Conclusion

Our analysis showed that beauty face filters will not significantly increase one's chance of being chosen as the more qualified candidate. On the other hand, ethnicity and experience have a strong impact on how people perceive job qualification. While some of our results are consistent with past research on hiring practices and biases within, we cannot add anything definitive to either argument of existing research on the effect of makeup or beauty transformations in the job market. Future studies could conduct similar experiments in a more realistic setting to obtain a more realistic representation of how beauty face filters affect one's chance of getting hired.

References

[1] Brown, D. (2020, April 1). 'your roots are showing': Photo editing apps surge after Salons shut down amid coronavirus. USA Today. Retrieved April 18, 2022, from https://www.usatoday.com/story/tech/2020/04/01/coronavirus-photo-video-editing-apps-surge/5087790002/

[2] Haines, A. (2021, December 10). From 'Instagram face' to 'snapchat dysmorphia': How beauty filters are changing the way we see ourselves. Forbes. Retrieved April 18, 2022, from https://www.forbes.com/sites/annahaines/2021/04/27/from-instagram-face-to-snapchat-dysmorphia-how-beauty-filters-are-changing-the-way-we-see-ourselves/?sh=455531db4eff

[3] Klatt, J., Eimler, S. C., & Krämer, N. C. (2016). Makeup your mind: The impact of styling on perceived competence and warmth of female leaders. The Journal of Social Psychology, 156(5), 483–497. https://doi.org/10.1080/00224545.2015.1129303

[4] Busetta, G., Fiorillo, F. & Palomba, G. The impact of attractiveness on job opportunities in Italy: a gender field experiment. Econ Polit 38, 171–201 (2021). https://doi.org/10.1007/s40888-020-00194-5

[5] Ford, R., Price, G., Hofmeyr, K. B., & Chiba, M. (2018). Brains versus beauty in the knowledge economy. SA Journal of Human Resource Management, 16. https://doi.org/10.4102/sajhrm.v16i0.897

[6] What sample sizes do you need for conjoint analysis? Q Research Software. (2020, November 24). Retrieved April 18, 2022, from https://www.qresearchsoftware.com/what-sample-sizes-do-you-need-for-conjoint-analysis

[7] Programming Choice Experiments in Qualtrics - m-graham.com. (n.d.). Retrieved April 19, 2022, from https://m-graham.com/resources/conjoint%20how-to.pdf

[8] Kline, P., Rose, E., & Walters, C. (2021). Systemic discrimination among large U.S. employers. https://doi.org/10.3386/w29053

Appendix

A. Unedited vs. Edited Photo Examples

Unedited | **Edited**

Figure 1: Unedited White Female

Figure 2: Edited White Female

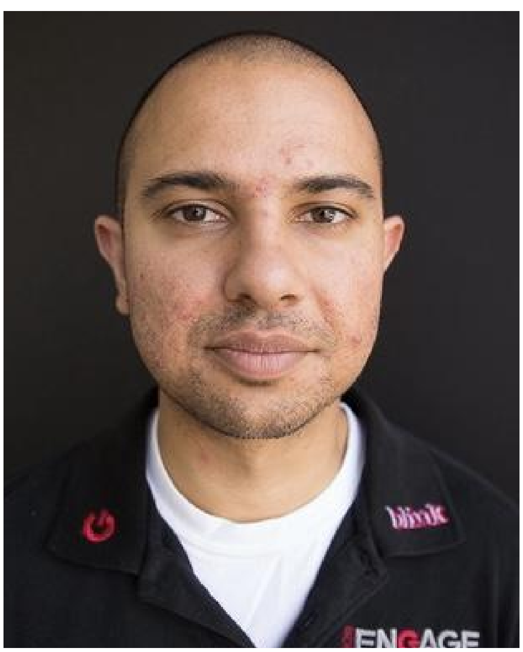


Figure 3: Unedited Non-White Male

Figure 4: Edited Non-White Male

B. Regression Results

Table 1: Full Data Regressions

	Dependent variable: Y		
	(1)	(2)	(3)
filtered	0.011	0.012	0.046*
	(0.021)	(0.016)	(0.027)
male		-0.040**	-0.013
		(0.016)	(0.022)
ethnicityWhite		0.069***	0.069***
		(0.016)	(0.016)
as.factor(exp)6		0.061***	0.060***
		(0.019)	(0.019)
as.factor(exp)7		0.101***	0.100***
		(0.019)	(0.019)
front_facing_role		-0.037**	-0.029
		(0.016)	(0.022)
top50		-0.014	-0.014
		(0.016)	(0.016)
filteredTRUE:male			-0.054*
			(0.031)
filteredTRUE:front_facing_role			-0.014
			(0.031)
Constant	0.494***	0.450***	0.434***
	(0.022)	(0.022)	(0.025)
Observations	4,070	4,070	4,070
R2	0.0001	0.015	0.015
Adjusted R2	-0.0001	0.013	0.013
Residual Std. Error	0.500 (df = 4068)	0.497 (df = 4062)	0.497 (df = 4060)
F Statistic	0.520 (df = 1; 4068)	8.667*** (df = 7; 4062)	7.093*** (df = 9; 4060)

Table 2: Subsetted Data: Females-only (1) vs. Males-only (2) Regressions

	Dependent variable:	
	Y	
	(1)	(2)
filtered	0.036	−0.019
	(0.029)	(0.031)
ethnicityWhite	0.025	0.151***
	(0.029)	(0.031)
as.factor(exp)6	0.072**	0.101***
	(0.036)	(0.036)
as.factor(exp)7	0.172***	0.139***
	(0.038)	(0.036)
front_facing_role	−0.063**	−0.107***
	(0.030)	(0.031)
top50	0.011	−0.034
	(0.030)	(0.029)
Constant	0.435***	0.400***
	(0.040)	(0.040)
Observations	1,127	1,035
R2	0.027	0.046
Adjusted R2	0.022	0.040
Residual Std. Error	0.494 (df = 1120)	0.490 (df = 1028)
F Statistic	5.143*** (df = 6; 1120)	8.234*** (df = 6; 1028)
Note:		*p<0.1; **p<0.05; ***p<0.01