

Changing looks and changing “discrimination:” The beauty of economists

Daniel S. Hamermesh^{*,1}

Abstract

Using candidates' multiple appearances in elections in a professional association with different photographs accompanying ballots, I show that exogenous increases in beauty raise chances of electoral success. The outcomes do not stem from correlations of ascriptive characteristics with unchanging unobserved differences in productivity-enhancing characteristics.

Keywords: Professional associations; Discrimination; Looks

JEL classification: J7

*Tel: +1-512-475-8526; fax +1-512-471-3510

E-mail address: hamermes@eco.utexas.edu (D.S. Hamermesh)

¹Edward Everett Hale Centennial professor of economics, University of Texas at Austin, research associate, National Bureau of Economic Research, and program director, Forschungsinstitut zur Zukunft der Arbeit (IZA).

1. Introduction

An immense empirical literature has produced estimates of differences in economic and other outcomes by race, gender, ethnicity, religion, sexual orientation and other ascriptive characteristics. Most of these studies have attempted to adjust for other (call them productivity-enhancing) characteristics that might affect the outcome of interest. The pervasive difficulty is that there may well be unobservable productivity-enhancing characteristics that are correlated with the ascriptive characteristic of interest, biasing estimates of the impact of the latter.

One approach to circumventing this correlation is audit studies—sending probes (resumés, individual testers) to random samples of decision-makers to infer how they react to agents who may be otherwise identical except for the ascriptive characteristic. (See Fix and Struyk, 1995, on race; Neumark *et al*, 1996, on gender; Weichselbaumer, 2003, on sexual orientation.) Their hypothetical nature and the difficulties in inferring what their findings mean (Heckman, 1998) mean that they fail to solve this correlation problem in the context of actual decision-making. Several studies (Blank, 1991; Goldin and Rouse, 2000; Straus *et al*, 2001) deal with this issue in real-world behavior by examining cases in which *all* the information on the ascriptive characteristic is removed. This solves one problem; but it is difficult to claim that the giant leap to no information about the characteristic will generate the same results as changing its distribution with no change in the average amount of information available to agents.

No study has examined how actual decisions are altered when individual variation in the information changes with no change in the average amount of information. We do that here in examining the effect of beauty, a characteristic that typically changes very

slowly and one that affects outcomes in labor markets and elsewhere through a variety of mechanisms (Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998; Mulford *et al*, 1998; Möbius and Rosenblat, 2006). We infer the impact of changed beauty under the assumption that no other objective characteristic or perception of the individual changes simultaneously.

2. Analytical Approach and Data

The outcome studied is election as an officer of the American Economic Association. Since 1935 the Association has conducted two annual four-person elections, for two slots as Vice-President and two as members of the Executive Committee. In addition to the ballot each AEA member receives an information sheet containing short biographies and 1-1/4" x 1-1/2" snapshots of each candidate. All of this information was available beginning with elections for officers for 1966 and continuing through the elections for 2004, i.e., a total of 312 candidacies. The 312 consist of only 216 different individuals, including 147 who appear once, 57 who appear twice, 13 who were nominated three times and 3 who appear four times. Multiple candidacies are crucial to the analysis. Because in most cases a candidate submitted a different picture each time s/he appeared on the ballot, the same candidate presented the voters with a different image in each candidacy. Even though the same person is on the ballot, the information provided to voters by the candidates' appearance may vary for this reason. Moreover, because his/her competition changes, even with the same picture a candidate's appearance relative to competitors can change.

The goal is to analyze whether a candidate's appearance affects his/her electoral chances and how an individual's changed appearance affects his/her chance of victory.²

We cannot determine whether voters react to the person's appearance as a signal of productivity, or whether voters have a preference for good looks independent of any signal. With over 4000 economists voting in each of the elections, however, only a small minority knows the candidates personally. Professional reputation, including that conveyed in the written material accompanying the ballot, may matter, as may the information contained in the photograph. By examining multiple candidacies we can, however, infer whether changing one's appearance affects how decision-makers (the voters) react to changes in the ascriptive characteristic beauty, holding constant other changes between ballots.³

As shown by Donald and Hamermesh (2007), the probability of victory in these elections is significantly positively affected by a candidate's scholarly impact (citations in journal articles, computed from the *Social Science Citation Index*)—his/her share of the citations received by the four candidates in the election; by being a woman, and by previously or currently holding a high-level government position. In addition to these measures we also include other variables in describing the outcome: Affiliation (Top 5 School, or nonacademic); race (African-American); field (theorist/econometrician); future Nobel Prize winner; and years since receiving the highest degree (essentially a proxy for age).

To examine the role of beauty in these elections we had the 312 photographs rated independently by each of four observers. We needed to make sure that the raters did not know any of the candidates, but that they were familiar with whatever *gestalt* academic economists may typically present.⁴ To deal with this difficulty I employed four entering first-year economics Ph.D. students, all of whom had attended undergraduate institutions

in the United States, as raters. To reflect the membership of the Association, three were male and one female. Each was asked to rate the attractiveness of the candidate in each picture on a scale from 10 (top) to 1 (bottom) and to try to maintain an average rating of 5 (without re-examining previously rated photos). I assume that the raters' perceptions represent the perceptions of the voters who confronted the photographs and the ballots.

Table 1 presents statistics describing the ratings. There are substantial differences among raters in the moments of the distribution of ratings. Accordingly, each rater i 's scores were normalized to yield standardized scores, the minima and maxima of which are shown in the final two columns of the first four rows. The six pair-wise correlations of these scores range from 0.28 to 0.40, with Cronbach's alpha equaling 0.65. These statistics indicate a very slightly lower concordance among the raters than has been observed in other studies (e.g., Pfann *et al*, 2000), possibly because the tiny photographs induce more noise than usual. Nonetheless, there is substantial agreement among raters: All the pair-wise correlation coefficients are highly significant statistically.

The four standardized ratings were averaged to obtain an average standardized rating, R_{ij} , for each candidate i in election j . A candidate's beauty relative to the entire history of candidates is unlikely to be relevant for his/her electoral chances. Rather, and analogous to our treatment of citations, it is his/her relative beauty in the election that may affect the outcome. Accordingly, for each candidacy we form:

$$(1) \quad R_{ij}^* = R_{ij} - R_j,$$

where R_j is the average of the average standardized beauty of all four candidates in the j 'th election. We focus on the impact of R_{ij}^* on the probability of election.

3. Effects of Beauty in Elections, All Candidacies 1966-2004

The fraction of winners among the half of the 312 candidacies with above-average beauty is 0.548 (standard error of mean = 0.040); among the half of the candidacies with below-average beauty the winning fraction is only 0.452 (standard error of mean = 0.040). This simple cut of the data by whether the candidate's beauty is above or below the average for his/her competitors suggests that beauty does matter in these elections. To examine this further, and as a baseline for comparing the results estimated over the multiple candidacies, Table 2 presents estimates of the impact of relative average standardized beauty on a candidate's chance of victory. Because there are two winners, standard probit/logit estimation methods cannot be used. While a conditional logit that used the six (${}_4C_2$) possible pairs of candidates in an election as observations would solve the problem of two winners, the set-up violates the assumption of the independence of irrelevant alternatives in the conditional logit (the absence of correlation of the unobservables across pairs). To deal with this difficulty the coefficients presented in Table 2 are based on the multinomial multiple response estimator developed in Donald and Hamermesh (2007). We assume that the errors follow the extreme value distribution, so that the index indicating a candidate's strength is estimated under assumptions analogous to those underlying ordinary logit estimation.⁵

The first column in Table 2 shows estimates of the basic equation determining the probability of election. Only the parameter estimates on citation share and female are presented: The estimated coefficients on most of the other variables—being in a Top 5 economics department, not being an academic, being a future Nobel, a theorist or econometrician, an African-American, and years of post-degree experience—had no

statistically or economically significant impacts. Having held a high-level government position has a statistically significant positive effect but is not of interest here. The second column adds R_{ij}^* to the equation. In this cross section the impact of beauty on the probability of being elected is positive with a t-statistic greater than one, although not statistically significant at conventional levels. The estimated effect of being better looking than the other candidates is not small: At one standard deviation below the mean beauty in an election an identical candidate has a chance of 0.44 of being elected; at one standard deviation above the mean the chance is 0.56.

Column (3) includes an interaction between R_{ij}^* and gender, based on previous work (Hamermesh and Biddle, 1994) that indicated that the impacts of beauty on outcomes are greater for males. The implications of the results are remarkable: In this sample a woman's looks relative to those of the other candidates in an election have absolutely no effect on her electoral chances. Among men, however, the impact is substantial and approaches statistical significance. Given the overriding importance of gender in these elections (74 percent of female candidates are elected), it is perhaps unsurprising that female candidates' beauty has little marginal effect.⁶

Columns (4) and (5) explore asymmetries in voters' responses to the candidates' relative beauty. The estimates in column (4) demonstrate that moving further below the average looks in an election has no impact on one's chances of winning; but it does pay to be increasingly better looking than one's competitors, and the effect is large and statistically significant. Moreover, as implied by the estimates in Column (5), the asymmetry only matters for male candidates. Being below or above average in looks has no impact on female candidates' victory probabilities.

4. Using Multiple Candidacies to Infer the Effects of Beauty

The cross-section results in Section 3 suggest that even economists pay attention to beauty when making decisions. For whatever reason this ascriptive characteristic matters in determining an outcome that we assume is the result of maximizing choices made by a set of presumably rational agents. The question is whether changes in the characteristic, holding constant other changes that might affect a candidate's electoral chances, alter the outcomes. To examine this possibility we estimate models over the 73 candidates (165 candidacies) who appear in two or more elections.

Variations in their perceived beauty are almost surely exogenous to the election outcomes: It is extremely difficult to argue that the candidates learn about the role of beauty from their prior electoral experiences and submit more appealing pictures on subsequent ballot appearances. Indeed, in their first electoral outing R_{ij} (R_{ij}^*) averaged 0.026 (0.003); in their final electoral outing, the second for most of the 73 sample members, R_{ij} (R_{ij}^*) averaged -0.083 (-0.010). The differences in average beauty between a candidate's first and final outings are statistically zero; but they do decrease, contrary to what one would expect if the candidates submitted pictures based on their learning about the role of beauty in the electoral process.

Columns (1) and (2) of Table 3 list estimates of the same equations as in Columns (2) and (4) of Table 2. (With the elimination of nearly half the sample and thus no adding-up constraint on predicted probabilities, ordinary logit estimation becomes appropriate.) The standard errors are robust, accounting for the clustering of observations. The parameter estimates and their implications are very similar to those estimated over the entire sample. The effect of R_{ij}^* is about the same (although the

statistical significance is lower). Again we observe an asymmetric impact of R_{ij}^* , with a statistically significant advantage to being increasingly better looking than the average candidate, but no disadvantage to being increasingly worse looking.

In Column (3) the same specification as in Column (1) is re-estimated over the reduced sample of 33 people who won at least one election and lost at least one. This sample is quite small, only 82 candidacies, so that sample size limitations lead us to expect a reduction in statistical significance compared to the earlier estimates. Despite that, the t-statistic exceeds one in absolute value; and the point estimate of the marginal impact of an increase in beauty in this sub-sample is larger than in the entire sample or in the sample of all multiple candidacies.

For comparison to other estimates, Column (4) presents logit estimates over this same sub-sample of candidacies, including only those variables that change between candidacies—citation share, relative average standardized beauty and years of post-degree experience.⁷ Again there is fairly little change in the estimated impact of beauty on electoral outcomes compared to the other specifications and (larger) samples. The parameter estimate is positive, and the t-statistic still exceeds one.

The conditional fixed-effect logit estimates are based on variations within the 33 candidates who appear in 82 elections and who lost at least one election and won at least one. The average duration between appearances on the ballot is 7.2 years, with a minimum of 1 year and a maximum of 21 years. In many cases these durations may be sufficient to allow the candidate's underlying beauty to change slightly.⁸ More important, however, with different photographs on the ballots voters' perceptions of the candidates' looks, as proxied by the raters' perceptions, will vary across a candidate's

appearances. The within-candidate variance of relative average standardized beauty is 39 percent of the total variance in this measure.

The estimates in the final column of Table 3 show the impacts of changes in candidates' citations and beauty relative to the other three candidates included in the elections in which they appear. All of the variation is within the candidate (relative to the changing variation among the competing candidates). Even with this stringent test we find roughly the same effect of changes in the candidate's relative beauty on his/her chances of winning an election as we have seen in the other estimates. Moreover, the standard error of the parameter estimate remains roughly what it was (conditional on the diminution of the sample size).⁹ The mean probability of election in this reduced sub-sample is 0.49. When R^*_{ij} drops by one standard deviation, the electoral probability falls to 0.43; when it increases by one standard deviation, it rises to 0.55. These effects are not small and are nearly identical to those in the cross-section estimates over the entire sample.¹⁰

5. Conclusions and Implications

This study has not shown that there is discrimination against bad-looking people. Rather, I have demonstrated that a particular real-world outcome becomes more favorable for the same person when perceptions of his/her looks improve exogenously. When sample sizes are reduced sufficiently, the estimated impacts of beauty are no longer significant statistically; but the estimated impacts of perceptions of beauty remain unchanged, and the uncertainty about the estimates increases solely in proportion to the decrease in sample sizes. These findings should strengthen the conclusion, both about beauty and by inference about other ascriptive characteristics such as race or gender, that

even with the same average amount of information it is relative position that is important. The characteristic itself, not any correlation with unobservable productivity-enhancing characteristics, generates the apparent impacts on outcomes of interest.

Acknowledgements

Stephen Trejo gave helpful comments on a previous draft, and Mark Pocock provided research assistance.

References

- Biddle, J., Hamermesh, D., 1998. Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre. *Journal of Labor Economics* 16, 172-201.
- Blank, R., 1991. The Effects of Double Blind versus Single Blind Reviewing: Experimental Evidence from the *American Economic Review*. *American Economic Review* 81, 1041-1067.
- Donald, S., Hamermesh, D., 2007. What Is Discrimination? Gender in the American Economic Association, 1935-2004. *American Economic Review*, 97, forthcoming.
- Mix, M., Struyk, R., 1995. Clear and Convincing Evidence: Measurement of Discrimination in America. Washington: Urban Institute Press.
- Goldin, C., Rouse, C., 2000. Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians. *American Economic Review* 90, 715-741.
- Hamermesh, D., Biddle, J., 1994. Beauty and the Labor Market. *American Economic Review* 84, 1174-1194.

- Hatfield, E., Sprecher, S., 1986. *Mirror, Mirror....* Albany, NY: State University of New York Press.
- Heckman, J., 1998. Detecting Discrimination. *Journal of Economic Perspectives* 12, 101-116.
- Klein, M., Rosar, U., 2005. Physische Attraktivität und Wahlerfolg: Eine empirische Analyse am Beispiel der Wahlkreiskandidaten bei der Bundestagswahl 2002 [Physical Attractiveness and Electoral Success: An Empirical Analysis of the Example of Electoral District Candidates in the 2002 Parliamentary Election. *Politische Vierteljahresschrift* 46, 266-290.
- Möbius, M., and Rosenblat, T., 2006. Why Beauty Matters. *American Economic Review* 96, forthcoming.
- Mulford, M., Orbell, J., Shatto, C., Stockard, J., 1998. Physical Attractiveness, Opportunity and Success in Everyday Exchange. *American Journal of Sociology* 103, 1565-1592.
- Neumark, D., Bank, R., van Nort, K., 1996. Sex Discrimination in Restaurant Hiring: An Audit Study. *Quarterly Journal of Economics* 111, 915-941.
- Pfann, G., Biddle, J., Bosman, C., Hamermesh, D., 2000. Business Success and Businesses' Beauty Capital. *Economics Letters* 67, 201-207.
- Straus, S., Miles, J., Levesque, L., 2001. The Effects of Videoconference, Telephone, and Face-To-Face Media on Interviewer and Applicant Judgments in Employment Interviews. *Journal of Management* 27, 363-381.
- Weichselbaumer, D., 2003. Sexual Orientation Discrimination in Hiring.. *Labour Economics* 10, 629-642.

Table 1. Beauty Evaluations, Individual and Composite (312 AEA Candidacies)

	Average	Std. Dev.	Standardized:	
			Minimum	Maximum
Individual Ratings:				
Male 1	5.43	1.21	-2.84	2.96
Male 2	5.20	1.89	-2.23	2.02
Male 3	4.71	1.20	-3.10	3.58
Female	5.95	1.34	-2.19	2.27
Average Standardized Rating (R_{ij}):	0	0.71	-1.80	2.71
Relative Average Standardized Rating (R_{ij}^*):	0	0.58	-1.73	2.00

Table 2. Multinomial Multiple Response Estimates of the Impacts of Several Independent Variables on the Probability of Election, Elections 1966-2004, (N = 312)*

Ind. Var.:	(1)	(2)	(3)	(4)	(5)
Citation Share	3.43 (0.79)	3.37 (0.81)	3.42 (0.78)	3.50 (0.82)	3.66 (0.85)
Female	1.59 (0.34)	1.52 (0.35)	1.57 (0.34)	1.49 (0.35)	1.97 (0.56)
Rel. Ave. Stdzd. Beauty		0.221 (0.186)	0.322 (0.212)		
Rel. Ave. Stdzd. Beauty > 0				0.936 (0.429)	1.330 (0.525)
Rel. Ave. Stdzd. Beauty <0				-0.399 (0.378)	-0.477 (0.414)
Female*Rel. Ave. Stdzd. Beauty			-0.428 (0.475)		
Female*Rel. Ave. Stdzd. Beauty > 0					-1.319 (0.837)
Female*Rel. Ave. Stdzd. Beauty <0					0.351 (1.135)
Log L	-112.81	-112.07	-111.67	-110.30	-109.00

*Standard errors in parentheses here and in Table 3. Also included in each equation are indicators of whether the candidate had held or currently holds a high-level government position, whether he/she was in a top-five economics department, whether he/she was not an academic, would eventually win a Nobel Prize, was a theorist or econometrician, was an African-American, and a continuous measure of years since Ph.D. (or other terminal degree).

Table 3. Logit and Conditional Logit Estimates of the Determinants of the Probability of Election, Multiple Candidacies in Elections 1966-2004*

	(1)	(2)	(3)	(4)	(5)
Ind. Var.:	Logit with robust standard errors				Cond. FE logit
Citation Share	5.24 (1.48)	5.35 (1.53)	3.67 (1.79)	2.55 (1.15)	13.63 (5.17)
Female	2.75 (0.75)	2.76 (0.80)	0.56 (1.07)		
Rel. Ave. Stdzd. Beauty	0.353 (0.382)		0.636 (0.542)	0.466 (0.415)	0.637 (0.598)
Rel. Ave. Stdzd. Beauty > 0		1.567 (0.689)			
Rel. Ave. Stdzd. Beauty < 0		-0.508 (0.549)			
Log L	-97.82	-95.92	-50.68	-51.69	-20.26
N (candidacies)	165	165	82	82	82
N (candidates)	73	73	33	33	33

*Also included in Columns (1) – (3) are indicators of whether the candidate had held or currently holds a high-level government position, where he/she was in a top-five economics department, whether he/she was not an academic, would eventually win a Nobel Prize, was a theorist or econometrician, was an African-American, and a continuous measure of years since Ph.D. (or other terminal degree). The logit in column (4), and the conditional logit in column (5) include only the variables listed and the measure of years since degree.

Footnotes

²To my knowledge Klein and Rosar (2005) is the only study even to consider one-time appearances of candidates in actual elections (rather than in hypothetical examples).

³Mythical anecdotal evidence that this is the case is provided by Eddie Murphy's "White Like Eddie" *Saturday Night Live* routine in the early 1980s.

⁴Several economists argued that this study would be impossible to conduct, since the beauty measure would be identically zero.

⁵Coefficients using standard logit techniques are very similar to the correct ones displayed in Table 2.

⁶In a reduced sample that excludes the 46 female candidacies the estimated impact of beauty among male candidates is larger than the estimates in Column (1) and is significant statistically.

⁷This last changes differentially across observations because of differences in the time intervals between candidates' appearances on the ballot.

⁸There is evidence, however, of extremely high autocorrelation in an individual's looks, even over many decades (Hatfield and Sprecher, 1986, pp. 282-283).

⁹The substantial increase in the estimated effect of scholarly productivity—the share of citations—is surprising. It does not stem from non-linearity in the impact of citations: In the cross-section estimates in Table 2 a nonlinear term in this measure is insignificantly different from zero. It may arise from the fact that within-candidate variation in the citation share is only 15 percent of the total variance.

¹⁰Re-estimating all the models in Table 3 by substituting R_{ij} for R_{ij}^* does not alter the conclusions about the importance of beauty. Indeed, the statistical significance of coefficients on R_{ij} is greater than that of the estimates in Table 3 in all specifications. This similarity is not surprising, as the within-candidate standard deviation of average standardized beauty accounts for 33 percent of the total variation in this measure.