Sex Roles
DOI 10.1007/s11199-016-0586-1

ORIGINAL ARTICLE

But You Don't Look Like A Scientist!: Women Scientists with Feminine Appearance are Deemed Less Likely to be Scientists

Sarah Banchefsky[1] · **Jacob Westfall**[2] · **Bernadette Park**[1] · **Charles M. Judd**[1]

Abstract Two studies examined whether subtle variations in feminine appearance erroneously convey a woman's likelihood of being a scientist. Eighty photos (half women) of tenured/tenure-track science, technology, engineering, and math (STEM) faculty at elite research universities were selected from the Internet. Participants, naïve to the targets' occupations, rated the photos on femininity and likelihood of being a scientist and an early childhood educator. Linear mixed model analysis treated both participants and stimuli as random factors, enabling generalization to other samples of participants and other samples of stimuli. Feminine appearance affected career judgments for female scientists (with increasing femininity decreasing the perceived likelihood of being a scientist and increasing the perceived likelihood of being an early childhood educator), but had no effect on judgments of male scientists. Study 2 replicated these findings with several key procedural modifications: the presentation of the stimuli was manipulated to either be blocked by gender or completely randomized, questions pertaining to the stimuli's appearance were removed, and a third career judgment likelihood rating was added to avoid tradeoffs between scientist and early childhood educator. In both studies, results suggest that for women pursuing STEM, feminine appearance may erroneously signal that they are not well suited for science.

Keywords Gendered appearance · Stereotypes · Femininity · Face perception · Physical appearance · Science · STEM · Sexism

In the summer of 2015, San Francisco based tech firm OneLogin featured photos of their own employees on advertising posters aimed at recruiting more engineers. One of the featured female employees, Isis Wenger, raised doubts about the campaign's veracity; apparently, some people found it improbable that this young woman could be an engineer simply because she did not look like one—she was far "too attractive" to be a "real engineer." In response to this criticism, the hashtag "iLookLikeAnEngineer" went viral on Twitter, with engineers of different ages, races, and genders posting their self-portraits in an effort to challenge notions of what engineers are "supposed to" look like (Zamon 2015).

Women remain disconcertingly underrepresented in STEM fields (science, technology, engineering, and math; National Science Foundation [NSF] 2015), in part due to differential gender roles (Eagly and Wood 2012; Eccles 1987), life goals (Ceci and Williams 2011; Diekman et al. 2010), and gender bias in STEM (Moss-Racusin et al. 2012). This research has focused almost exclusively on categorical gender gaps between women's and men's experiences and outcomes in STEM. The present research, on the other hand, examines how bias might vary within gender categories. Specifically, we test whether real, accomplished female scientists judged as more feminine in appearance are also deemed less likely to be scientists.

The perceived incompatibility between femininity and science is a recognized issue with negative consequences for women. Over a 5-year period, 80 % of female and 72 % of male undergraduate engineering majors surveyed agreed that the belief that women in science or technical fields are

✉ Sarah Banchefsky
sarah.banchefsky@colorado.edu

[1] Department of Psychology and Neuroscience, University of Colorado Boulder, Muenzinger D244, Boulder, CO 80309-0345, USA

[2] Department of Psychology, The University of Texas at Austin, SEA 4.208, 108 E. Dean Keeton Stop A8000, Austin, TX 78712-1043, USA

Published online: 05 February 2016

unfeminine is a problem for women pursuing these careers; indeed, the more that a woman perceived that this was a problem, the less satisfied she was in her field (Hartman and Hartman 2008). Women in STEM environments have reported feeling unable to present themselves in a stereotypically feminine manner (e.g., wearing a skirt, expressing emotions) because they do not want to draw attention to their gender or are apprehensive that they will seem unsuitable for a STEM career (Hewlett et al. 2008; Pronin et al. 2004).

Indeed, appearance has a powerful and immediate effect on person perception. In just fractions of a second, people come to remarkably similar conclusions about a person's attractiveness, aggression, likability, trustworthiness, and competence (Willis and Todorov 2006). Like the major social categories of gender and race (Ito and Urland 2003), variations in facial appearance are automatically processed and automatically activate stereotypes. For example, stronger stereotypic Black features (e.g., broad nose, thick lips) activate Black stereotypes (e.g., close with family, failing in school; Blair et al. 2002, 2004a). Similarly, baby-faced features (e.g., large round eyes and foreheads) activate youthful characteristics (e.g., naïve, submissive; Zebrowitz et al. 1991). These stereotypes have potentially dramatic implications. For example, statistically controlling for the target's perceived categorical race (i.e., White or Black) and the seriousness of their crimes, convicted felons with more Afrocentric features receive harsher criminal sentences (Blair et al. 2004a), and baby-faced individuals are seen as less suitable for leadership jobs (Zebrowitz et al. 1991).

Research in non-STEM domains has examined the impact of femininity on judgments. For example, more feminine-appearing women were (accurately) judged as more likely to be Republicans than Democrats (Carpinella and Johnson 2013); moreover, female politicians with increasingly masculine facial appearance were less likely to receive votes, particularly among conservative constituents (Hehman et al. 2014). This work suggests that, at least in the political realm, women are rewarded for looking more feminine. The present research examines whether naturalistic variations in feminine appearance (i.e., based on physiological characteristics such as facial bone structure, as well as gender performance such as the use of make-up, hair style, etc.) impacts the perceived likelihood that a woman is a scientist. We hypothesized that women judged as increasingly feminine in appearance will also be judged as less likely to be scientists. We did not have a strong rationale from which to hypothesize about male targets. On the one hand, feminine-appearing men may similarly activate feminine gender stereotypes, thereby decreasing their judged likelihood of being scientists. On the other hand, they may also trigger "nerdy" male stereotypes (Cheng 2008) that align with stereotypes about the types of people who populate STEM domains (Cheryan et al. 2011), thereby increasing their judged likelihood of being scientists.

To rigorously test this hypothesis, we attended to the conceptual and methodological issues of stimulus sampling (Wells and Windschitl 1999) by employing a large stimulus set and treating the stimuli as random in the analysis (Judd et al. 2012). In traditional analyses, only participants (or more rarely, stimuli) are treated as the sole random factor in the analyses—variation across responses due to the other factor are averaged and thus ignored. In contrast, the cross-random model used in the present studies takes into account that we have two different samples from two theoretical populations of interest about which we would like to make inferences—a sample of participants and a sample of stimuli.

Most of the prior research on the effects of femininity has employed a very small sample of stimuli (e.g., one woman dressed in a feminine manner or a neutral manner; Betz and Sekaquaptewa 2012; six photographs, Sczesny and Kühnen 2004) that are sometimes not naturalistic (e.g., computer-generated; Friedman and Zebrowitz 1992), raising the question of whether results are simply due to the specific stimuli presented. Moreover, in part due to small samples of stimuli, previous studies have failed to examine the continuum of gendered appearance (i.e., from masculine to feminine), instead only focusing on extreme examples of masculinity and femininity on either end of the spectrum. This limited sampling not only constricts power and the generalizability of the effects but also is not a realistic representation of real people encountered on a daily basis.

In contrast, the present methodological approach offers several strengths and advantages. First, we treated variations in gendered appearance continuously rather than operationalizing or analyzing gendered appearance in a categorical way (see Irwin and McClelland 2003). Second, we used a large sample of photographs of real people, specifically tenured or tenure-track faculty members in STEM departments at elite U.S. universities. Third, we treated both stimuli (i.e., faces) and participants as random factors (Baayen et al. 2008; Judd et al. 2012). Critically, our analysis permits generalization from our specific sample of faces to other samples of faces that we might have used (Clark 1973; Judd et al. 2012).

Study 1

Method

Participants

Participants were 51 U.S.-based workers on Amazon.com's Mechanical Turk (25 men, 26 women; 78 % White, 12 % Asian, 4 % biracial, 4 % Latino, and 2 % Black; Mean age = 34.92, *SD* = 13.71, range = 18–63 years old) who were compensated \$0.75 for their time. An additional four participants failed two or more of four basic attention checks that were embedded within the survey (e.g., "Is this person's hair blonde or brunette?") and were excluded from the sample. An approximate power analysis based on the calculations given by Westfall et al. (2014), assuming a counterbalanced design and using their default variance partitioning coefficients, suggests that an experiment with 80 stimuli and 50 participants should have 80 % power to detect an effect size as small as Cohen's $d = .33$.

Stimuli

Stimuli consisted of 80 photographs (40 men, 40 women) of tenured/tenure-track faculty in elite STEM departments in U.S. universities. Programs were selected according to U.S. News and World Report's rankings of premier graduate programs in various STEM disciplines. Our stimulus selection rule was to select from each program's website the first high-quality, crisp, color photos of faculty who were smiling and making direct eye contact with the camera (some websites did not present photos). Moreover, in order to avoid variations in judged likelihood of being a scientist due to perceived race, all faculty selected appeared to be White. To ensure the faces were naturalistic, they were not cropped (e.g., to remove hair) and were presented in color. Example stimuli are available from the authors upon request. In total, nine research universities (Massachusetts Institute of Technology, California Institute of Technology, Princeton University, Stanford University, University of Texas at Austin, University of California Berkeley, University of Illinois, Cornell University, and Carnegie-Mellon University), and 15 STEM programs (aerospace engineering, astronomy, astrophysics, bioengineering, chemical engineering, chemistry, civil engineering, computer engineering, computer science, electrical engineering, engineering, environmental engineering, math, mathematics, and physics) were represented.

Although the majority of photographs selected were simply the first encountered that fulfilled the preceding criteria, about ten faces of each gender were strategically selected by the researchers to maximize representation along the spectrum from masculine to feminine appearance. That is, about five highly feminine- and five highly masculine-appearing women and men (relative to the rest of the stimulus sample) were chosen based on a holistic first impression of gendered appearance. We made these selections because research demonstrates that interaction effects (which we were hypothesizing) can be difficult to detect without adequate variation in the continuous variable of interest (i.e., gendered appearance); we therefore made an effort to select some faces that clearly varied in gendered appearance (McClelland and Judd 1993). Nevertheless, the variation was naturalistic and the selected individuals very much resemble typical people encountered in everyday life. Specifically, subjectively masculine-appearing individuals tended to have shorter hair and stronger facial features (e.g., heavier jaws and brow-bones, larger noses), whereas subjectively feminine-appearing individuals tended to have longer hair and finer facial features (e.g., smaller jaws and brow-bones). Some women wore jewelry (e.g., earrings, a necklace) or subtle make-up (e.g., faint lipstick). For both genders, although there was variation in clothing, no articles were revealing or flashy; nearly all targets wore solid-colored sweaters, tee-shirts or button-down shirts, or shirts with a subtle pattern. Some men and women appeared to have on a blazer, and some men wore a tie. Some of both genders wore eyeglasses. Finally, for both genders, some photos were taken inside offices or against a blank background, whereas others were taken outdoors.

Importantly, the statistical model employed to analyze the data (a mixed model treating both participants and stimuli as random factors, described in the following) took into account idiosyncratic differences among the stimuli, accounting for this naturalistic variation. Put simply, the analysis enabled us to detect whether judgments of femininity and career likelihood were related over and above any unique variations between stimuli, eliminating the possibility that arbitrary variations among the photographs gave rise to the observed effects (e.g., the relationship obtained between perceived femininity and career likelihood; Judd et al. 2012).

Procedure

Participants were asked to evaluate 80 photographs of individuals who, as described previously, were in fact accomplished academic STEM scientists. Participants did not know the targets' occupations, however, and were simply told that the study was about first impressions, and that first impressions are made very quickly and are often surprisingly accurate. They were then asked to rate each photo on three 7-point scales ranging from 1 (*not at all*) to 7 (*very*): masculine to feminine, likable to unlikable, and unattractive to attractive, in this fixed order. Note that although gendered appearance was measured on a semantic differential scale from masculine to feminine, we frequently refer to this variable as "feminine appearance" because femininity is the primary construct of interest.

Next, participants estimated the likelihood that the individual was a scientist, followed by the likelihood that the person was an early childhood educator (henceforth referred to as "teacher," a profession that is 97 % women and stereotypically feminine; Carnevale et al. 2013), on 6-point scales ranging from 1 (*very unlikely*) to 6 (*very likely*). Finally, they estimated the age of the target, selecting one of eight 5-year ranges starting at 25 years-old and ending at 60 and above. Participants had as long as they desired to make the ratings of each face before moving on. The target gender of the stimuli was blocked and counterbalanced (i.e., all women were presented first or all men were presented first). Photos within each block were randomized, and each was presented on a separate screen. Lastly, participants completed a series of demographic questions.

Data Analysis

Mixed models were estimated in SAS® using Satterthwaite approximate degrees of freedom. Initial analyses examined whether participant's gender or target's gender order (i.e., whether participants rated men or women first) affected any of the judgments. No effects were present so these variables are dropped from subsequent analyses. To examine whether gendered appearance impacted career ratings, career likelihood was analyzed as a function of career (science vs. teacher, contrast coded), face gender (male vs. female, contrast coded), judged feminine appearance (mean-centered), and all possible interactions. Data were analyzed using linear mixed models with crossed random effects of participants and stimuli (meaning that both participants and stimuli were treated as random effects and every participant rated every stimulus; Baayen et al. 2008; Judd et al. 2012). All possible random intercepts and slopes were estimated (Barr et al. 2013), but the covariances between the random effects were not estimated due to convergence problems (estimating covariances involved estimating an additional 81 parameters).

In addition to using mixed models in which both participants and stimuli were treated as random effects, Study 1 entailed another statistical advantage: because our experimental design involved every participant providing judgments of gendered appearance for every face, this variable varied both between and within participants, as well as between and within faces. It was therefore possible to decompose gendered appearance ratings into three distinct effects based on three different sources of variance: the target effect (i.e., judgments due to the face), the perceiver effect (i.e., judgments due to the perceiver), and the relationship effect (i.e., judgments due to the face and the perceiver). These effects are similar by analogy to effects estimated under the Social Relations Model (Kenny 1994), which our analysis resembles. For the descriptions that follow, let F_{ij} denote the femininity rating given by the i^{th} participant to the j^{th} face.

> **Case 1: Target**. The target effect is the average level of femininity that a given face elicits across all perceivers. It asks whether certain faces are evaluated as more or less feminine on average, relative to other faces, and how this deviance affects judgments of career likelihood. It is computed as $\bar{F}_j$ and then mean-centered in the mixed model.
> **Case 2: Perceiver**. The perceiver effect represents a participant's average rating tendency for femininity. It asks whether certain participants on average evaluated faces as more or less feminine, relative to other participants, and how this deviance affects judgments of career likelihood. It is computed as $\bar{F}_i$ and then mean-centered in the mixed model.
> **Case 3: Relationship**. The relationship effect (i.e., Target × Perceiver interaction) examines a perceiver's rating of a particular face, asking how much it deviates from the face's average femininity rating and the perceiver's average femininity rating tendency. It asks whether a given participant perceives a given face as more or less feminine than would be expected and how this affects judgments of career likelihood. It is computed as $F_{ij}-\bar{F}_j-\bar{F}_i$ and then mean-centered in the mixed model (see Raudenbush 2009; Rosnow and Rosenthal 1991).

Including each of these predictors allows an appraisal of how femininity ratings due to stimuli, perceivers, and the relationship between them each uniquely predicts judgments of career likelihood. Importantly, decomposing the femininity predictor in this way also avoids the problem of biased parameter estimates that can result from pooling together effects from different levels of analysis (Bafumi and Gelman 2006; Bell and Jones 2015).

We predicted that the three-way interaction of interest (Target Gender × Career Type × Femininity) would emerge for two of these effects: first, for the target effect, this interaction would indicate that women who are judged as more feminine than others, on average, are judged as less likely to be scientists relative to teachers (Case 1, target effect); second, an interaction involving femininity due to relationship would indicate that when a given participant views a particular woman as more feminine (over and above the participant's typical femininity ratings, as well as the face's average femininity rating), he or she also views that woman as less likely to be a scientist relative to a teacher (Case 3, relationship effect); the third possible effect (Case 2, perceiver effect) would mean that those perceivers who on average see greater femininity across all faces also judged all of the women on average as less likely to be scientists. Although we did not hypothesize this effect, we included it in the model so as to avoid biased parameter estimates.

Results

Preliminary Analyses

In preliminary analyses that treated face as the unit of analysis and averaged across participants, the male and female scientists were perceived as about the same age, equally likable (the unlikability rating was reverse-scored for interpretability in Table 1), and equally attractive (see Table 1). Unsurprisingly, female scientists were rated as significantly more feminine in appearance than male scientists, and they also were rated as

Table 1 Mean ratings and correlations by face gender on six face dimensions in study 1

Feature dimension	Mean (SD)		Correlations					
	Female faces	Male faces	1.	2.	3.	4.	5.	6.
1. Feminine	5.11 (1.03)$_a$	2.85 (.78)$_b$	–	.89**	.59**	−.41**	−.56**	.75**
2. Attractive	4.25 (.92)$_a$	4.04 (.63)$_a$	−.48**	–	.69**	−.63**	−.61**	.67**
3. Likeable	3.94 (.40)$_a$	3.92 (.44)$_a$	.28	.68**	–	−.45**	−.25	.65**
4. Age	3.11 (1.16)$_a$	2.80 (1.16)$_a$	−.69**	−.08	−.11	–	.31*	−.31*
5. Likelihood scientist	3.90 (.43)$_a$	3.96 (.53)$_a$	−.13	−.29	.11	.50**	–	−.65**
6. Likelihood teacher	3.93 (.37)$_a$	3.14 (.34)$_b$	−.01	.54**	.76**	−.37*	−.25	–

SD Standard deviation. Means with different subscripts are significantly different, $p < .05$. Age ratings were categorical and represented ranges in years (1 = 25–29; 2 = 30–35; 3 = 36–40; 4 = 41–45; 5 = 46–50; 6 = 51–55; 7 = 56–60; and 8 = 61+). Correlations are based on averages for each face. Correlations for female faces are above the diagonal. Correlations for male faces are below the diagonal. *$p < .05$; **$p < .01$

significantly more likely to be teachers. Notably, female and male scientists were rated as equally likely to be scientists (see Table 1). As depicted in Table 1, average judged feminine appearance and attractiveness were highly positively correlated for female scientists and less strongly but negatively correlated for male scientists. Femininity was also negatively correlated with age, but more so for male than female scientists. Due to the high correlation between femininity and attractiveness for female scientists (.89), and collinearity problems this creates in the predictors, we did not include both simultaneously in analyses, instead analyzing their effects in separate models. Ancillary analyses reported subsequently examined age and femininity simultaneously to confirm that perceived femininity affected career judgments over and above perceived age.

Feminine Appearance and Career Likelihood

Table 2 presents fixed effects output (effect sizes in the form of unstandardized beta estimates are also presented. Other effect size estimates are not presented because there is no generally agreed upon definition of standardized effect size estimates for mixed models; Snijders and Bosker 1994). First, the hypothesized three-way interaction among Target Gender, Career Type, and Feminine Appearance was significant (target effect), $F(1, 97.9) = 6.10$, $p = .015$. We broke down this interaction by target gender, examining female and male scientists separately. As hypothesized, feminine appearance affected career judgments for female scientists (Career Type × Femininity), $F(1, 90.7) = 25.00$, $p < .001$, but had no impact on career judgments for male scientists ($p = .53$). Consistent with our hypothesis, as the average rated feminine appearance of a female scientist increased, she was judged as significantly less likely to be a scientist, $F(1, 78.3) = 12.67$, $p < .001$ (Fig. 1a), and significantly more likely to be a teacher, $F(1, 74.3) = 41.99$, $p < .001$ (Fig. 1b). This pattern is shown by the bold-type regression lines in each panel of Fig. 1, where the slopes in the two panels for female scientists (but not male scientists) are different from each other and are both significantly different from zero.

A marginal three-way interaction emerged among Target Gender, Career Type, and Feminine Appearance (relationship effect), $F(1, 48.5) = 3.96$, $p = .052$. As before, the relationship effect of feminine appearance affected career judgments for female scientists (Career Type × Femininity), $F(1, 66.2) = 12.60$, $p < .001$, but had no impact on career judgments for male scientists ($p = .80$). This pattern is shown by the thin, shorter regression lines in each panel of Fig. 1, which depict the within-stimulus regressions of career likelihood on femininity (with perceiver effects removed) for each face. Breaking this interaction down further by career indicated that the interaction for female scientists was driven by teacher judgments: when a given participant viewed a given woman as having more feminine appearance than expected (based on the perceiver's typical femininity rating and the face's typical femininity rating), he or she also rated her as more likely to be a teacher, $F(1, 5918) = 38.81$, $p < .001$ (Fig. 1b); however, perceiving a given face as more feminine in appearance than expected (for that participant and for that target) did not affect the perceived likelihood of being a scientist ($p = .97$, Fig. 1a). Thus the thin regression lines in Fig. 1 show a significant positive slope on average only for teacher ratings of the female scientists.

There were no significant effects attributable to perceiver differences in judged feminine appearance. Lower order effects emerged in the model, but all were qualified by the two reported three-way interactions (see Table 2).

Table 2 Mixed-models results for fixed effects for career likelihood judgments in study 1

Effect	Estimate	*SE*	*df*	*t*	*p*
Intercept	3.697	.075	81.3	49.090	< .0001
Main effects					
Career	.331	.070	112	4.690	< .0001
Gender	.192	.046	95.3	4.180	< .0001
Feminine_Face (Target effect)	−.011	.030	78.4	−.370	.715
Feminine_Ss (Perceiver effect)	.145	.116	47.6	1.250	.218
Feminine_Rel (Relationship effect)	.032	.013	43.5	2.380	.022
Two-way interactions					
Career × Gender	−.054	.061	95.3	−.870	.385
Career × Feminine_Face (Target effect)	−.146	.043	93.3	−3.410	.001
Career × Feminine_Ss (Perceiver effect)	−.121	.081	45.7	−1.490	.142
Career × Feminine_Rel (Relationship effect)	−.047	.017	59.6	−2.830	.006
Gender × Feminine_Face (Target effect)	.031	.032	85.8	.960	.339
Gender × Feminine_Ss (Perceiver effect)	−.049	.040	46.6	−1.220	.228
Gender × Feminine_Rel (Relationship effect)	.031	.013	51.7	2.270	.028
Three-way interactions					
Career × Gender × Feminine_Face (Target effect)[a]	−.109	.044	97.9	−2.470	.015
Career × Gender × Feminine_Ss (Perceiver effect)	.064	.052	40	1.240	.221
Career × Gender × Feminine_Rel (Relationship effect)[a]	−.042	.021	48.5	−1.990	.053

Estimate unstandardized beta, *SE* standard error, *df* Satterthwaite approximate degrees of freedom

[a] Indicates hypothesized effect of interest

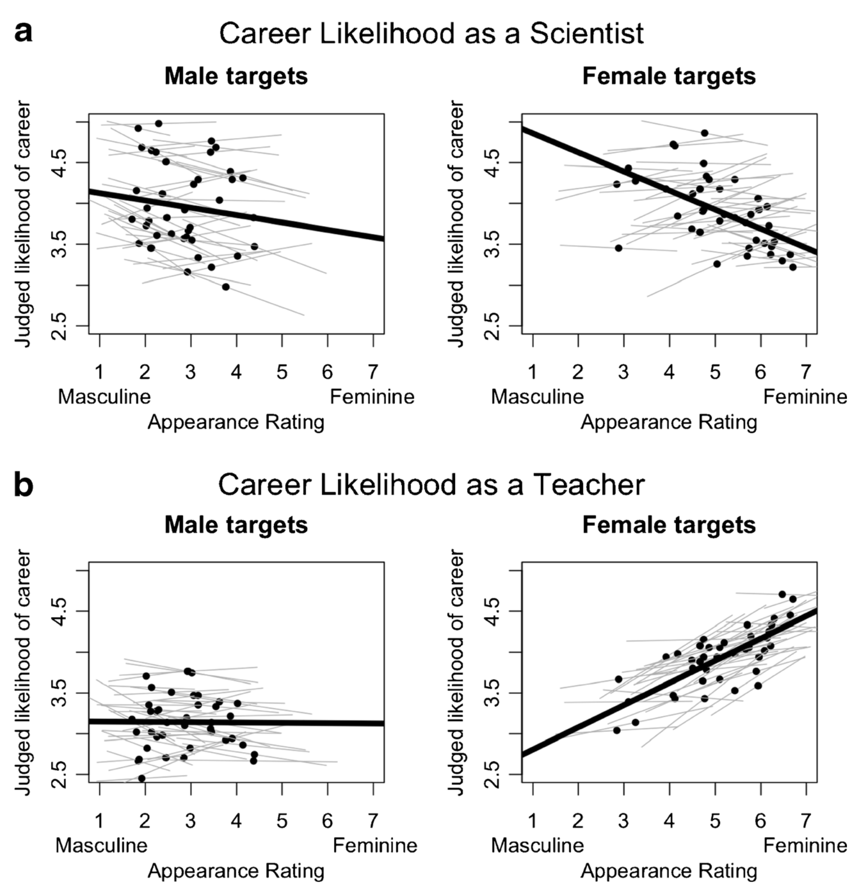


Fig. 1 Plot of mixed model results by target gender and career. The points in each panel represent the mean femininity and likelihood ratings for each stimulus face (i.e., the target effects), and the bold regression line in each panel of career likelihood ratings on femininity ratings represents the total target effect in that panel. The thin, shorter regression lines passing through each target effect represent the within-stimulus regressions of career likelihood on femininity (with perceiver effects removed; i.e., the relationship effects), the average of which represents the total relationship effect in that panel. For both scientist and teacher judgments, the bold lines are significantly different from 0 for female targets, but not for male targets

Ancillary Analyses

Because age was correlated with feminine appearance, and because we were concerned that perhaps younger looking individuals might be viewed as less likely to be scientists than older ones, we also examined a model that controlled for the target's perceived age (again parsed into the same three sources of variation and mean-centered). There were no significant effects of perceived age on career judgments, and the critical three-way interactions involving target gender, career type, and feminine appearance remained unchanged when controlling for age in the model.

The target's mean attractiveness and feminine appearance were very highly correlated (.89) for female scientists. Because of potential collinearity problems, we decided against running models that included both as simultaneous predictors. Instead, we estimated a separate model identical to that estimated for feminine appearance but using attractiveness judgments instead of gendered appearance judgments. This model revealed significant two-way Career Type × Attractiveness interactions (again for both target effect and relationship effect), indicating that more attractive scientists were seen as less likely to be scientists and more likely to be teachers (target effect: $F(1, 105) = 28.09$, $p < .001$; relationship effect: $F(1, 55.7) = 10.86$, $p = .002$). Unlike the effects of femininity, these effects did not depend on target gender. These findings align with research suggesting that people who pursue science are stereotyped as unattractive (Hannover and Kessels 2004). They also suggest that attractiveness is used as a cue for judging both men and women's career likelihood, whereas only for women is gendered appearance also used as an informative career-likelihood cue.

Study 2

One concern in Study 1 is that asking participants to evaluate the targets' appearance (e.g., attractiveness, femininity) may have made these concepts especially salient or created pressure to be consistent in how one related appearance judgments and career likelihood judgments. Another potential concern addressed in Study 2 is that the blocked presentation of the stimuli (by target gender) produced excessive attention to within-category variations in appearance. We hypothesized that femininity would still be used as a cue to career-type even when between-category differences in gender were made more salient by presenting male and female targets interspersed (see Blair et al. 2002). Finally, we were concerned that having participants rate only two careers may have forced participants to make a trade-off in judging career likelihood that they would not have made if more careers had been assessed. Study 2 addressed these concerns in addition to testing the replicability of the effects in a larger sample.

Method

Participants

Because a between-subjects condition factor was added, a larger sample of 214 people participated in the study on Amazon's Mechanical Turk (129 women, 84 men; approximately 80 % White, 6 % Black, 4 % Latino, 5 % Asian, 4 % Biracial, and 1 % Native American; Mean age = 36.27, $SD = 11.41$, range = 18–68 years old). An approximate power analysis using the same assumptions as in Study 1 (with a slight adjustment reflecting the fact that participants in Study 2 did not make gendered appearance judgments) indicated that these sample sizes should provide 80 % power to detect effect sizes as small as $d = .37$ (Westfall et al. 2014). The survey took about 20–30 min to complete and workers were paid $.75. No attention checks were included, so no participants were excluded from analysis.

Design

Study 2 replicated Study 1 with the following alterations: First, participants were randomly assigned to judge faces presented in either a blocked or mixed fashion with respect to target gender. In the blocked condition, participants rated either all women followed by all men or vice versa, with the order of the target gender blocks counterbalanced, and faces were presented in a randomized order within each block (as in Study 1). In the mixed condition, all faces were presented in a fully randomized order for each participant, theoretically making the gender of the target more salient. Second, participants only made career-likelihood judgments of each target, ensuring that explicit considerations of femininity or other appearance-related measures would not influence career-likelihood judgments. Third, to make it less apparent that we were examining a male stereotypic (scientist) and female stereotypic career (teacher), participants first rated a relatively gender-neutral career, journalist (64 % female; Carnevale et al. 2013), for each target. Even in this more conservative design, we hypothesized that perceivers would still use female targets' feminine appearance as a career cue.

Procedure

The cover story was very similar to Study 1. Participants were randomly assigned to judge faces either blocked by gender ($n = 103$) or in a mixed presentation ($n = 111$). Participants rated each face in terms of their likelihood of being a journalist, scientist, and early childhood educator (teacher), in that order, and again on 6-point scales from 1(*very unlikely*) to 6 (*very likely*). Participants lastly completed the same demographics as in Study 1.

Data Analysis

Career likelihood was analyzed as a function of career type (scientist vs. non-scientist, and teacher vs. journalist, two single degree of freedom contrasts), target gender (male vs. female, contrast coded), gendered appearance (mean-centered), presentation (blocked vs. mixed, contrast coded), participant gender (male vs. female, contrast coded), and all possible interactions. The gendered appearance variable was based on the average femininity rating for each face in Study 1. We hypothesized that the three-way interaction of interest (Target Gender × Career Type × Feminine Appearance) would again emerge, and that although it might vary in strength as a function of mixed versus blocked presentation, it would be significant in both conditions. Data were again analyzed using linear mixed models with crossed random effects of participants and stimuli; as before, all possible random intercepts and slopes were estimated, but not covariances. Because perceivers in Study 2 did not make gendered appearance judgments, all femininity effects reported for Study 2 are "target effects" (i.e., "perceiver" and "relationship" effects could not be estimated).

Results

Feminine Appearance and Career Likelihood

Table 3 presents the fixed effects of our hypothesized interactions. First, the predicted significant three-way interaction among Target Gender, Career Type (Science vs. Other), and Feminine Appearance was found again, $F(1, 81.4) = 26.94$, $p < .001$. As can be seen in Fig. 2, whereas feminine appearance again affected career judgments for female scientists (Science vs. Non-Scientist × Femininity), $F(1, 78.5) = 43.16$, $p < .001$, it had no impact on career judgments for male scientists (i.e., this two-way interaction was not significant for men, $p = .11$). Consistent with our hypothesis, the feminine appearance of female targets was negatively related to perceived likelihood of being a scientist, $F(1, 76.8) = 26.83$, $p < .001$, and positively related to perceived likelihood of being a non-scientist, $F(1, 77.7) = 58.37$, $p < .001$. The lack of a Teacher vs. Journalist × Target Gender × Feminine Appearance interaction indicated that femininity affected career likelihood judgments of journalist the same way that it affected ratings of teacher.

The impact of feminine appearance was also moderated by participant gender, $F(1, 201) = 6.15$, $p = .01$; although the critical three-way interaction was highly significant for both male and female participants, it was stronger among female participants ($F(1, 86.4) = 32.49$, $p < .001$) than among male participants, $F(1, 94.8) = 19.54$, $p < .001$. If anything, this suggests that women perceivers especially may consider another woman's gendered appearance as a meaningful cue of her career. However, because participants' gender did not moderate the results in Study 1, this finding should be interpreted with caution. A variety of unanticipated lower order effects emerged, but all were importantly moderated by the predicted three-way interaction (see Tables 4 and 5).

Interestingly, the intermixed presentation of male and female stimuli enhanced categorical gender bias in career judgments (regardless of feminine appearance). That is, a significant Target Gender × Presentation × Science vs. Other

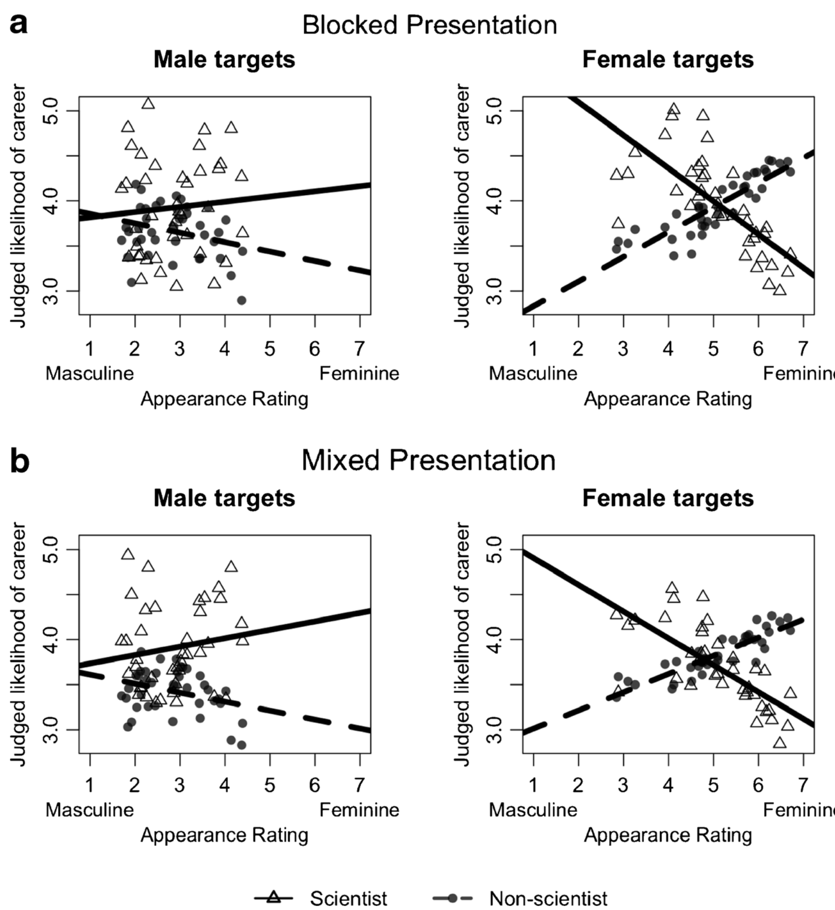
Table 3 Mixed-models results for predictors of interest fixed effects for career likelihood judgments in study 2

Effect	Estimate	*SE*	df	*t*	*p*
Intercept	3.534	.052	229	68.00	<.0001
Main effects (Predictors of interest)					
Scientist vs. Non-scientist (Scientist)	.278	.053	100	5.23	<.0001
Teacher vs. Journalist (T vs. J)	−.108	.057	115	−1.91	.058
Target gender	.101	.037	91.1	2.70	.008
Femininity	.064	.025	77.5	2.54	.013
Two-way interactions (Predictors of interest)					
Scientist × Target gender	.278	.050	81.2	−.19	.849
Scientist × Femininity	−.108	.035	78.0	−2.68	.009
T vs. J × Target gender	.101	.053	91.9	4.38	<.0001
T vs. J × Femininity	.064	.036	82.5	1.05	.295
Target gender × Femininity	.278	.026	82.7	6.49	<.0001
Three-way interactions (Predictors of interest)					
Scientist × Target gender × Femininity[a]	−.184	.035	81.4	−5.19	<.0001
T vs. J × Target gender × Femininity	−.052	.037	86.2	−1.42	.159

T vs. *J* Teacher versus Journalist, *Estimate* unstandardized beta, *SE* standard error; *df* Satterthwaite approximate degrees of freedom

[a] Indicates hypothesized effect of interest

Fig. 2 Plot of mixed model results by target gender, career, and presentation type. For female targets, all slopes are significantly different from 0, whereas for male targets, all slopes are statistically equivalent to 0



interaction indicated that differences in judged career likelihood for male vs. female targets was stronger in the mixed condition than in the blocked condition, $F(1, 226) = 22.28$, $p < .001$ (see Table 4). Simple effects looking within target revealed that male targets were judged as more likely to be scientists (compared to other careers) in the mixed vs. blocked presentation, $F(1, 287) = 6.25$, $p = .013$, whereas female targets were judged as less likely to be scientists (compared to other careers) in the mixed vs. blocked presentation, $F(1, 271) = 8.58$, $p = .004$.

Ancillary Analyses

When age was included as a predictor (again using average age ratings for each face from Study 1), the critical three-way interaction among target, feminine appearance and career remained highly significant. We also examined a model that included perceived attractiveness rather than feminine appearance (again using average attractiveness ratings for each face from Study 1); consistent with Study 1, a Career Type × Attractiveness interaction indicated that more attractive targets were seen as less likely to be scientists and more likely to be non-scientists, $F(1, 99.6) = 94.28$, $p < .001$. This did not depend on target gender, suggesting that attractiveness affected career judgments similarly for male and female targets.

Discussion

Two studies examined how variation in judged gendered appearance of 80 real scientists related to judgments about their likelihood of being a scientist. Participants were unaware that the photographs they were judging were actually scientists; rather, they were simply told that they were making first impressions of individuals. Results showed that for female scientists, but not male scientists, perceivers used gendered appearance as a cue about how likely they were to be scientists (vs. early childhood educators/teachers or journalists). Study 2 demonstrated that this outcome was the case (a) regardless of whether male and female scientists were presented in a blocked or intermixed order; (b) when participants were not asked to judge the person's appearance prior to making career judgments (i.e., when aspects of appearance were not made salient); and (c) even when an additional, gender-neutral career (journalist) was included in the career judgments alongside scientist and early childhood educator. In both studies, these results did not depend on participant gender.

Table 4 Fixed effects results for presentation (blocked by gender versus unblocked) on career likelihood judgments, study 2

Presentation effects (Mixed vs. Blocked)	Estimate	*SE*	df	*t*	*p*
Presentation	−.067	.039	222	−1.72	.087
Presentation × Scientist	−.003	.022	245	−.16	.875
Presentation × T vs. J	−.039	.028	224	−1.4	.161
Presentation × Target gender	.053	.015	380	3.52	.001
Presentation × Femininity	−.016	.007	511	−2.08	.038
Presentation × Participant gender	.004	.039	222	.09	.927
Presentation × Scientist × Target gender	−.066	.014	226	−4.72	<.0001
Presentation × Scientist × Femininity	.020	.009	167	2.32	.021
Presentation × Scientist × Participant gender	−.017	.021	230	−.78	.433
Presentation × T vs. J × Target gender	.069	.020	207	3.39	.001
Presentation × T vs. J × Femininity	−.004	.011	145	−.37	.713
Presentation × T vs. J × Participant gender	.048	.027	217	1.74	.084
Presentation × Target gender × Femininity	−.019	.009	279	−2.1	.037
Presentation × Target gender × Participant gender	−.019	.015	380	−1.27	.206
Presentation × Femininity × Participant gender	−.002	.007	511	−.21	.835
[a]Presentation × Scientist × Target gender × Femininity	.017	.010	201	1.71	.090
Presentation × Scientist × Target gender × Participant gender	−.013	.013	241	−.99	.324
Presentation × Scientist × Femininity × Participant gender	.009	.008	188	1.16	.248
Presentation × T vs. J × Target gender × Femininity	−.007	.013	167	−.55	.582
Presentation × T vs. J × Target gender × Participant gender	.012	.020	204	.6	.550
Presentation × T vs. J × Femininity × Participant gender	−.006	.011	145	−.56	.577
Presentation × Target gender × Femininity × Participant gender	.000	.009	279	.03	.972
Presentation × Scientist × Target gender × Femininity × Participant gender	−.015	.009	203	−1.58	.116
Presentation × T vs. J × Target gender × Femininity × Participant gender	.004	.012	164	.36	.718

T vs. *J* Teacher versus Journalist, *Estimate* unstandardized beta, *SE* standard error, *df* Satterthwaite approximate degrees of freedom

[a] Highest order effects of interest. Other effects of interest are in bold

Overall, the female scientist's gendered appearance was related to judgments about the likelihood of being in a masculine-stereotypic career (science), a feminine-stereotypic career (teacher), and even a relatively gender-

Table 5 Remaining fixed effects results for gender on career likelihood judgments, study 2

Effect	Estimate	*SE*	*df*	t	*p*
Participant gender	−.011	.039	222	−.28	.781
Participant gender × Scientist	.037	.022	234	1.72	.087
Participant gender × T vs. J	−.025	.028	223	−.89	.375
Participant gender × Target gender	.000	.015	380	0	.996
Participant gender × Femininity	.016	.007	511	2.16	.031
Participant gender × Scientist × Target gender	.030	.013	234	2.26	.025
Participant gender × Science × Femininity	−.022	.008	178	−2.84	.005
Participant gender × T vs. J × Target gender	.003	.020	204	.17	.865
Participant gender × T vs. J × Femininity	.025	.011	140	2.19	.030
Participant gender × Target gender × Femininity	.013	.009	279	1.49	.138
[a]Participant gender × Scientist × Target gender × Femininity	−.023	.009	201	−2.48	.014
Participant gender × T vs. J × Target gender × Femininity	.013	.013	163	1.07	.286

T vs. *J* Teacher versus Journalist, *SE* standard error, *df* Satterthwaite approximate degrees of freedom

[a] Highest order effect of interest

neutral career, journalism. On the other hand, a male scientists' gendered appearance was not related to career likelihood regardless of career type. In short, a woman's gendered appearance was used as a cue about her career in a way that a man's gendered appearance was not. In contrast to gendered appearance, men and women alike who were judged as more attractive were deemed less likely to be scientists and more likely to be non-scientists. Aligning with previous research, feminine appearance and attractiveness were strongly positively correlated for female targets, and less strongly but negatively correlated for male targets (Perrett et al. 1998).

The methodology and statistical approach of the present research has several important advantages. In Study 1, every participant evaluated all 80 faces, allowing an examination of three different sources of variation in femininity ratings—ratings due to targets, perceivers, and the relationship between the two. A good deal of research neglects to examine divergent sources of variance, which can mask important relationships in the data (Bell and Jones 2015; Kievit et al. 2013). Finally, the use of mixed models with crossed random effects ensures that the results are not simply an artifact of the specific stimuli selected for our study. Rather, the stimuli were treated as a random factor—that is, as just one possible sample of stimuli drawn with error from the population of interest. Theoretically, if we were to conduct the study again with a different stimulus set of top scientists, our results suggest that we should expect similar estimates (Judd et al. 2012).

Although past research has suggested that femininity and attractiveness are generally viewed as incompatible with science (Hartman and Hartman 2008; Pronin et al. 2004), this is the first research we know of to use a naturalistic, robust stimulus set and demonstrate that subtle variations in gendered appearance alter perceptions that a given woman is a scientist. This research employed a more robust and sophisticated stimulus set and analytic approach relative to previous research regarding the negative implications of femininity for women in the workplace (Sczesny and Kühnen 2004). Specifically, past research has relied on non-naturalistic (e.g., computer-generated, hand-drawn) stimuli, and/or a very minimal number of stimuli (e.g., using the same female "job applicant" dressed in a sexy vs. non-sexy manner; see Glick et al. 2005; using the same female role model and having her dress in feminine or non-feminine clothing; Betz and Sekaquaptewa 2012). In these studies, the generalizability of the findings is limited because the results could simply be due to the specific (and potentially extreme) stimuli selected (Wells and Windschitl 1999).

By using a large number of naturalistic photographs (of scientists) and appropriately treating stimulus as a random factor in the analysis (i.e., as just one selection of stimuli from the theoretical population of interest), the present studies have a number of strengths. First, given that all targets were indeed scientists, we can rule out that participants were responding to a real relationship between appearance and the likelihood of being a scientist. Second, gendered appearance varied along a continuum—as it does in real life—rather than only representing extremes. Third, our statistical analysis supports the idea that idiosyncratic differences between the individual photographs did not give rise to the results. Finally, the statistical analysis supports the idea that a different selection of photographs should theoretically obtain the same results.

Practice Implications

Our work has a number of implications. First, we would recommend that scientists who are already established within STEM fields strive to celebrate and highlight existing diversity within STEM—both between social categories (e.g., different genders or racial groups) but also within social categories (see Galinsky et al. 2015). People are drawn to fields where they feel they would belong and be similar to others (Hannover and Kessels 2004). In addition to being discouraged by male-dominated STEM environments (Murphy et al. 2007) or those populated by male stereotypic objects (Cheryan et al. 2009), women's interest in STEM may also be thwarted by the undue perception that women scientists cannot express femininity. The #iLookLikeAnEngineer campaign exemplifies a marketing strategy that challenges stereotypes about what engineers look like. Already, this hashtag is being touted by several companies and universities to display the variety of individuals within engineering. Given that exposure to counter-stereotypic STEM role models has been shown to increase men and women's interest in STEM, such a strategy should benefit men and women and boys and girls alike (Cheryan et al. 2012).

Such campaigns may also alleviate pressure on women in STEM to suppress their femininity. Indeed, research has found that some women in STEM not only minimize feminine appearance (e.g., avoid wearing make-up) but also eschew feminine traits, behaviors, and goals (e.g., being emotional, leaving work to raise children; Pronin et al. 2004). Problematically, cultures that devalue femininity can also lead women to distance themselves from and criticize other women (Ellemers et al. 2004), especially feminine women (Rhoton 2011). Such practices reinforce the perceived incompatibility between femininity and STEM, bolster the status quo, minimize the diversity that women have to offer to STEM fields, and are harmful to the women involved—leading to isolation, dissatisfaction and potential abandonment with their field (Hartman and Hartman 2008; Hewlett et al. 2008).

To counteract pressure to assimilate and become "one of the guys," people in male-dominated or male stereotypic fields should strive to cultivate an environment that celebrates diversity and where individuals feel as though they can present themselves in whatever way they choose (Galinsky et al. 2015). Indeed, research shows that racial minorities in companies felt more engaged in their work ("I am proud to tell others I work [for this organization]") to the extent that their White colleagues endorsed a multicultural perspective that celebrated and recognized diversity (e.g., "Employees should recognize and celebrate racial and ethnic differences") rather than an assimilationist perspective that maintained minorities should strive to be more like Whites (e.g., "Employees should downplay their racial and ethnic differences"; Plaut et al. 2009, pp. 444). Ideologies about how best to approach group differences also exist concerning gender (Hahn et al. 2015). Given that such ideological perspectives are malleable (Wolsko et al. 2000) and shift between various workplace environments (Plaut et al. 2009), STEM environments might aim to embrace an ideological approach that is more welcoming to women and people who generally do not fit the stereotypical STEM mold. Indeed, when racial minorities rated their interest in a company where they would be numerically underrepresented (compared to one where they would be more equally represented), they preferred a company with a multicultural perspective rather than a colorblind perspective. This is likely because a colorblind company, coupled with numeric underrepresentation of one's social group, implied that their racial identity was not valued or welcomed and that they would be expected to minimize their social identity and assimilate to the predominant group (Purdie-Vaughns et al. 2008).

At the very least, we should not conclude that feminine or "girly images" of women in STEM are uniformly harmful to fostering women's interest in STEM (Betz 2012; Betz and Sekaquaptewa 2012). In our opinion, there is an important distinction between portraying naturalistic variation in women's gendered appearance in STEM versus extreme, objectified or sexualized portrayals of feminine women scientists. The latter approach to fostering women's interest in STEM has proven to be ineffective. For example, the European Commission launched a campaign entitled "Science, it's a girl thing" in an effort to convey that science can be feminine. However, their promotional video, which was ultimately withdrawn due to criticism, featured young women strutting around the lab in high-heels and mini-skirts, playing with make-up and blowing kisses at test-tubes as a male scientist observed them (Khazan 2012). Although such depictions of feminine women in science are clearly problematic, it remains unclear how exposure to naturalistic variation in gendered appearance in STEM might cultivate greater interest in STEM fields. It is our hope that future research will further explore the conditions under which feminine appearing women in STEM inspire and motivate others.

Limitations and Future Directions

Despite the strengths of our research, several questions remain to be addressed in the future. One clear limitation of the present research is that we intentionally used only White men and women scientists to avoid the possibility of arousing intersecting race or ethnic biases. For example, common race-based stereotypes maintain that Asians are better at math than Whites (Aronson et al. 1999) and that Blacks are less academically capable than Whites (Steele and Aronson 1995). Such stereotypes may have affected the perceived likelihood of being a scientist for both male and female targets, and in potentially different ways (e.g., for Black men vs. women; Shields 2008). Because our primary question concerned differential use of gendered appearance of femininity as a career cue for male vs. female targets, we presented individuals who all appeared to be White. That said, future research should examine whether our findings extend to people of other apparent races and ethnicities.

Another interesting and important issue is precisely which aspects of appearance participants were using to make judgments about femininity (e.g., inherent facial structure or facial features vs. performed femininity such as hairstyle and make-up), and how each of these might differentially contribute to inferences, attributions, and career judgments about a target. For example, performed femininity, such as wearing make-up, in contrast to femininity in facial structure, may be viewed as particularly incompatible with STEM careers because it suggests that a woman puts too much effort or time into her appearance.

Some readers may wonder whether participants were picking up on a real phenomenon whereby women in STEM are really less feminine than other women. This asks an empirical question that remains to be addressed—might women in STEM actually be objectively less feminine in appearance on average (see Carpinella and Johnson 2013)? Although future research is needed to examine this issue (and to understand why this may or may not matter), the present research demonstrates that regardless of whether this is the case, there is certainly variation in gendered appearance among scientists, and this variation is used as a cue of a woman's—but not a man's—likelihood of being a scientist. Whether or not women in STEM are less feminine-appearing on average than women outside STEM, there clearly is variation in appearance within STEM fields in terms of gendered appearance and attractiveness, and this variation is sufficient to elicit bias (see Hewlett et al. 2008; Seymour and Hewitt 1997).

Moreover, although career likelihood may seem like a somewhat innocuous outcome compared to, for example, decisions to hire an individual, it a) represents a validated phenomenon wherein feminine women in STEM fields report encountering doubt about their likelihood of being a member of a STEM field (e.g., "But you don't look like a programmer!"; Hewlett et al. 2008; Zamon 2015) and b) is likely related to potentially more blatant forms of bias. For example, role incongruity theory maintains that perceiving that a person does not "fit" within a career (i.e., that they are unlikely to be in that career) due to a mismatch between that person's gender role and the career role elicits prejudice towards the individual (Eagly and Karau 2002; Heilman 2012). Thus, a woman who is more feminine in appearance than other women will elicit stronger perceived role incongruity and will therefore experience more prejudice than their less feminine counterparts (Eagly and Karau 2002; Heilman 2012). Future research is required to validate other forms of bias that feminine women in STEM might encounter.

Another lingering question is the extent to which using femininity in social judgments occurs automatically and, relatedly, whether it could be controlled. The robust correlation found between feminine appearance and career likelihood is particularly disconcerting because unlike sensitivity to potential categorical biases (e.g., being sexist or racist), people tend to be less aware and capable of controlling biases based on within-category variations, despite being clearly informed about how such biases operate and asked not to use them (Blair et al. 2004b; Sczesny and Kühnen 2004). This suggests that even when evaluating only women for a position or conscientiously combating gender bias, feminine women may nevertheless evoke more negative judgments. Indeed, research regarding Afrocentricity suggests that providing clear diagnostic information about an individual still did not override the influence of within-category appearance cues on judgments about that individual (Blair et al. 2005). Future research is warranted to examine whether people are aware that they are using feminine appearance in making judgments and whether they can overcome such responses.

Finally, a further exploration of how these processes affect women is warranted. What happens to a woman when she is explicitly told, or signaled in some way, that it does not look as though she belongs in a given field? How do such interactions affect women across their lifetime? For example, before choosing science, are feminine girls and women—because they don't "look" like scientists—treated differently by parents, teachers, and others (Tenenbaum and Leaper 2003)? What about relatively masculine girls and women who pursue stereotypically feminine careers—do they encounter additional hurdles simply based on their appearance? Such interactions may elicit a cascade of inferences that not only guide the perceiver's behavior, but in turn affect the self-perceptions and behavior of the girls and women themselves (Snyder et al. 1977).

Conclusions

The present paper opened with a story about Isis Wenger, a woman whose legitimacy as a computer engineer was contested when her photograph was featured in a recruiting advertisement for her company. Our results suggest that her story is not an isolated event—in our studies, men and women alike used women's gendered appearance, but not men's, as an indication that they were less likely to be scientists (and more likely to be teachers and journalists). This work empirically validates claims made by some women in STEM that their belonging or aptitude in their career has been doubted simply due to their feminine appearance, and it contributes to research suggesting that appearance is more valued, scrutinized, and consequential for women than men (Bar-Tal and Saxe 1976; Feingold 1990; Hehman et al. 2014). Documenting what may be a novel type of gender bias, the present work indicated that gendered appearance was uniquely used as a cue to a women's career but not a man's career. Overall, our findings suggest that for women, within-category variation in feminine appearance has the potential to negatively impact the current national strategic goal of creating a diverse, welcoming, and egalitarian STEM workforce.

Compliance with Ethical Standards The University of Colorado Boulder's Institutional Review Board (IRB) approved the research presented in the manuscript. This research did not receive external funding.

Conflict of Interest There were no conflicts of interest in conducting this research.

References

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29–46.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bafumi, J., & Gelman, A. E. (2006). *Fitting multilevel models when predictors and group effects correlate*. Retrieved from http://academiccommons.columbia.edu/item/ac:125243

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bar-Tal, D., & Saxe, L. (1976). Physical attractiveness and its relationship to sex-role stereotyping. *Sex Roles, 2*(2), 123–133.

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods, 3*(1), 133–153.

Betz, D. (2012, September 11). The trouble with Barbie science. *Scientific American.* Retrieved from scientificamerican.com.

Betz, D. E., & Sekaquaptewa, D. (2012). My fair physicist? Feminine math and science role models demotivate young girls. *Social Psychological and Personality Science, 3*(6), 738–746.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology, 83*(1), 5–25.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004a). The influence of Afrocentric facial features in criminal sentence. *Psychological Science, 15*(10), 674–679.

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004b). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology, 87*(6), 763–778.

Blair, I. V., Chapleau, K. M., & Judd, C. M. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology, 35*(1), 59–68.

Carnevale, A. P., Strohl, J., & Melton, M. (2013). What's it worth?: The economic value of college majors. *Georgetown University Institutional Report*. Retrieved from https://cew.georgetown.edu/wp-content/uploads/2014/11/whatsitworth-complete.pdf

Carpinella, C. M., & Johnson, K. L. (2013). Politics of the face: The role of sex-typicality in trait assessments of politicians. *Social Cognition, 31*(6), 770–779.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences, 108*(8), 3157–3162.

Cheng, C. (2008). Marginalized masculinities and hegemonic masculinity: An introduction. *The Journal of Men's Studies, 7*(3), 295–315.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*(6), 1045–1060.

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? *Social Psychological and Personality Science, 2*(6), 656–664.

Cheryan, S., Drury, B. J., & Vichayapai, M. (2012). Enduring influence of stereotypical computer science role models on women's academic aspirations. *Psychology of Women Quarterly, 37*(1), 72–79.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335–359.

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science, 21*(8), 1051–1057.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573–598.

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology* (pp. 458–476). Thousand Oaks: Sage Publications.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*(2), 135–172.

Ellemers, N., Van den Heuvel, H., Gilder, D., Maass, A., & Bonvini, A. (2004). The underrepresentation of women in science: Differential commitment or the queen bee syndrome? *British Journal of Social Psychology, 43*(3), 315–338.

Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology, 59*(5), 981–993.

Friedman, H., & Zebrowitz, L. A. (1992). The contribution of typical sex differences in facial maturity to sex role stereotypes. *Personality and Social Psychology Bulletin, 18*(4), 430–438.

Galinsky, A. D., Todd, A. R., Homan, A. C., Phillips, K. W., Apfelbaum, E. P., Sasaki, S. J., … Maddux, W. W. (2015). Maximizing the gains and minimizing the pains of diversity: A policy perspective. *Perspectives on Psychological Science, 10*(6), 742–748.

Glick, P., Larsen, S., Johnson, C., & Branstiter, H. (2005). Evaluations of sexy women in low and high status jobs. *Psychology of Women Quarterly, 29*(4), 389–395.

Hahn, A., Banchefsky, S., Park, B., & Judd, C. M. (2015). Measuring intergroup ideologies positive and negative aspects of emphasizing versus looking beyond group differences. *Personality and Social Psychology Bulletin, 41*(12), 1646–1664.

Hannover, B., & Kessels, U. (2004). Self-to-prototype matching as a strategy for making academic choices: Why high school students do not like math and science. *Learning and Instruction, 14*, 51–67.

Hartman, H., & Hartman, M. (2008). How undergraduate engineering students perceive women's (and men's) problems in science, math and engineering. *Sex Roles, 58*, 251–265.

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science, 5*(7), 815–824.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior, 32*, 113–135.

Hewlett, S., Luce, C., Servon, L., Sherbin, L., Shiller, P., Sosnovich, E., … Sumberg, K (2008). *The Athena factor: Reversing the brain drain in science, engineering, and technology.* Watertown: Harvard Business School.

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research, 40*(3), 366–371.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*(4), 616–626.

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*(1), 54–69.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Khazan, O. (2012, June 22). E.U.'s 'Science, it's a girl thing' campaign sparks a backlash. *The Washington Post.* Retrieved from www.washingtonpost.com

Kievit, R., Frankenhuis, W. E., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Quantitative Psychology and Measurement, 4*, 1–14.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*(2), 376–390.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474–16479.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*(10), 879–885.

National Science Foundation, National Center for Science and Engineering Statistics. (2015). *Women, minorities, and persons with disabilities in science and engineering: 2015*. Special Report NSF 15–311. Arlington, VA. Retrieved from http://www.nsf.gov/statistics/wmpd/

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., … Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394*(6696), 884–887.

Plaut, V. C., Thomas, K. M., & Goren, M. J. (2009). Is multiculturalism or color blindness better for minorities? *Psychological Science, 20*(4), 444–446.

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology, 40*(2), 152–168.

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology, 94*(4), 615–630.

Raudenbush, S. W. (2009). Adaptive centering with random effects: An alternative to the fixed effects model for studying time-varying treatments in school settings. *Education Finance and Policy, 4*(4), 468–491.

Rhoton, L. A. (2011). Distancing as a gendered barrier understanding women scientists' gender practices. *Gender and Society, 25*(6), 696–716.

Rosnow, R. L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at *only* the interaction (unless all main effects are zero). *Psychological Bulletin, 110*(3), 574–576.

Sczesny, S., & Kühnen, U. (2004). Meta-cognition about biological sex and gender-stereotypic physical appearance: Consequences for the assessment of leadership competence. *Personality and Social Psychology Bulletin, 30*(1), 13–21.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder: Westview.

Shields, S. A. (2008). Gender: An intersectionality perspective. *Sex Roles, 59*(5), 301–311.

Snijders, T. A., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research, 22*(3), 342–363.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*(9), 656–666.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797–811.

Tenenbaum, R., & Leaper, C. (2003). Parent–child conversations about science: The socialization of gender inequities? *Developmental Psychology, 39*(1), 34–47.

Wells, G., & Windschitl, P. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*(9), 1115–1125.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020–2045.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598.

Wolsko, C., Park, B., Judd, C. M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology, 78*(4), 635–654.

Zamon, R. (2015, August 4). *#ILookLikeAnEngineer reminds us that anyone can (and should) be an engineer.* Retrieved from http://www.huffingtonpost.ca/2015/08/04/ilooklikeanengineer_n_7934098.html

Zebrowitz, L. A., Tenenbaum, D. R., & Goldstein, L. H. (1991). The impact of job applicants' facial maturity, gender, and academic achievement on hiring recommendations. *Journal of Applied Social Psychology, 21*(7), 525–548.