THE ROLE OF PERCEIVER GENDER IDEOLOGY AND TARGET FEMININITY IN IMPLICIT AND EXPLICIT GENDER-SCIENCE STEREOTYPES

By

SARAH MARY BANCHEFSKY

B.A., MIAMI UNIVERSITY, 2009

M.A., UNIVERSITY OF COLORADO BOULDER, 2011

A thesis submitted to the

Faculty of the graduate school of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2015

This thesis entitled:
The Role of Perceiver Gender Ideology and Target Femininity in Implicit and Explicit Gender-Science Stereotypes
written by Sarah Mary Banchefsky
has been approved for the Department of Psychology and Neuroscience

(Dr. Bernadette Park, Chair)

(Dr. Charles Judd)

(Dr. Irene Blair)

(Dr. Lawrence Williams)

(Dr. Randall O'Reilly)

June 30, 2015

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

HRC protocol 13-0419

ABSTRACT

Banchefsky, Sarah Mary (Ph.D., Psychology and Neuroscience)

The Role of Perceiver Gender Ideology and Target Femininity in Implicit and Explicit Gender-Science Stereotypes

Thesis directed by Professor Bernadette Park

The present research examined a previously unexamined stereotype—that men possess more scientific traits than women—and two moderators of this gender-science stereotype: sociopolitical beliefs about how to approach gender differences (gender ideology) and target feminine facial appearance (gender prototypicality). A pre-test to select face stimuli first established that female scientists with more feminine facial appearance were judged as less likely to be scientists (Study 1). A second pre-test validated both positive and negative traits along the scientific and warmth dimension that were viewed as stereotypic of scientists. Study 2 found that science stereotypic traits were more readily associated with men than with women. Specifically, men were more associated with scientific and cold traits, and less associated with unscientific and warm traits, relative to women, and these stereotypes emerged whether assessed at an implicit level (using a go/no-go association task; GNAT) or an explicit level (using a percent estimate task), but were uncorrelated with one another. Moreover, gender ideology moderated explicit stereotypes but had no effect on implicit stereotypes; but greater assimilationism (i.e., believing that women should be more like men in the workplace) and segregationism (i.e., believing that men and women belong in separate roles) both predicted stronger gender-science stereotypes, whereas greater gender blindness (i.e., believing each person should be treated as a unique individual regardless of his or her gender) predicted weaker stereotypes. To examine whether gender prototypicality influenced the strength of implicit stereotypes, three versions of the GNAT instantiated gender using either generic names (control) or photos of high prototypical or low prototypical men and women (pre-tested in Study 1). Despite evidence that women's facial prototypicality influenced perceived likelihood of being a scientist in Study 1, variation in prototypicality had very little impact on implicit gender-science stereotypes in Study 2. Other correlates of the gender ideologies are examined, and the implications, limitations and future directions of this work are considered.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: General Introduction

Biochemist and Nobel laureate Dr. Tim Hunt recently spoke at a world conference of science journalists in Seoul, Korea. During his invited speech at a luncheon sponsored by female scientists and engineers, Hunt said, “Let me tell you about my trouble with girls. Three things happen when they are in the lab. You fall in love with them, they fall in love with you, and when you criticize them, they cry.” He went on to say that he did not want to impede women’s success in science, but that he simply prefers single-sex labs. Criticism spread rapidly throughout social media, and within days, Hunt was asked to resign as Honorary Professor at the University of College London, and from the European Research Council (ERC), where Hunt served on a science committee that he helped establish (Chappell, 2015).

Such comments are likely to be particularly harmful in male-dominated fields such as STEM. Although women constitute the majority among recent college and university graduates and are rapidly entering fields once exclusive to men (e.g., law, medicine, and business), they remain underrepresented in college majors in the physical sciences, technology, engineering, and math (STEM), majors that open doors to highly-solicited and lucrative careers (Carnevale, Strohl, & Melton, 2013). Men outnumber women in practically every field of engineering and science, particularly physics, engineering, and computer science, where women obtain only 20 percent of bachelor’s degrees (Hill, Corbett, & Rose, 2010). In fact, the percentage of bachelor’s degrees in computer science awarded to women in the United States actually decreased from 28% to 18% between 2000 and 2010. Those in engineering decreased from 21% to 18%, and in mathematics/statistics they decreased from 48% to 43% (National Science Foundation, 2015). Although women currently claim the majority of advanced degrees in a variety of prestigious areas (52% of PhDs in life sciences, 57% in social sciences, 71% in psychology, and 77% of veterinary degrees; Hill et al., 2010), they compose only 8.8-15.8% of tenure-track positions in math-intensive fields at the top 100 universities in the United States (Ceci & Williams, 2011).

As a sizeable body of literature accumulates regarding the gender disparity in STEM (e.g., for reviews, see Ceci & Williams, 2011; Ceci, Williams, & Barnett, 2009; Halpern et al., 2007; Spelke, 2005), it indicates that the gender gap is likely not due to inherent differences in math ability (Halpern et al., 2007) but to pervasive gender stereotypes that underlie discrimination (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012), stereotype threat (Spencer, Steele, & Quinn, 1999), women's dampened sense of belonging and self-efficacy in STEM (Good, Rattan, & Dweck, 2012; Rhoton, 2011), and women's greater desire to help others and have a family (Ceci & Williams, 2011, 2012; Evans & Diekman, 2009; Frome, Alfeld, Eccles, & Barber, 2006; Weisgram, Bigler, & Liben, 2010; Weisgram, Dinella, & Fulcher, 2011)—goals perceived as especially unattainable in STEM careers (Cook & Minnotte, 2008; Diekman, Brown, Johnston, & Clark, 2010; Diekman, Clark, Johnston, Brown, & Steinberg, 2011).

Research on gender stereotypes has shown that men are more strongly associated with math and science than women, and that people tend to agree that men are better at, or more interested in, math and science than women (e.g., Kiefer & Sekaquaptewa, 2007; Nosek et al., 2009; Nosek & Smyth, 2011; Schmader, Johns, & Barquissau, 2004; Thoman, White, Yamawaki, & Koishi, 2008). Surprisingly however, this research has not examined whether women are perceived as lacking the *traits* typical among scientists. Although gender traits have been extensively examined in empirical research, this work has focused almost singularly on characteristics related to warmth and competence (or communality and agency), showing that women are ascribed more warm/communal traits (e.g., nurturing, caring, compassionate) than men, and men are ascribed more competent/agentic traits (e.g., aggressive, ambitious, dominant) than women (Cejka & Eagly, 1999; Dasgupta & Asgari, 2004; Eagly & Karau, 2002; Fiske, Glick, & Xu, 2002; Rudman & Kilianski, 2000; Williams & Best, 1990).

Importantly, stereotypes about scientists (e.g., nerdy, isolated, obsessed with technology) seem quite distinct from traditional agentic traits (e.g., competitive, stands up under pressure).[1] First, the traditional competence dimension lacks traits regarding scientific rigor, such as “scientific” and “rational.” Second, the “nerd” stereotype of scientists constitutes a subtype of masculinity that is actually defined in part by the absence of power or status seeking traits, and the presence instead of somewhat feminine characteristics such as lack of sports ability, small body size, physically weak, and lack of sexual relationships with women (Kendall, 1999; Smiler, 2006). Moreover, communal and agentic characteristics do not capture the social isolation and awkwardness typically associated with scientists. Whether women are perceived as less scientific than men therefore remains an open question.

In accordance with social role theory (Eagly, 1987), I hypothesize that women will be judged as lacking scientific traits because they are underrepresented in scientific roles. A first step in examining this hypothesis was validating a set of both positive and negative traits seen as stereotypic of scientists. Assessing the overlap between scientific traits and gender traits is important because if women are stereotyped as possessing fewer scientific traits than men, they will not only be seen as having less potential for science (due to descriptive stereotypes, e.g., “women are not scientific”), but will be evaluated more negatively when actually enacting scientist behaviors (due to prescriptive stereotypes, e.g., “women should not be scientific”; Burgess & Borgida, 1999). Indeed, research shows that women are uniquely devalued in leadership roles and other male-stereotypic roles because they are perceived as lacking the characteristics deemed necessary to succeed in such roles (Eagly & Karau, 2002; Glick, Zion, & Nelson, 1988; Koenig et al., 2011; Swim, Borgida, Maruyama, & Myers, 1989; see also Heilman, 1983).

In addition to establishing this gender-science stereotype, the present research examines two potential moderators of the stereotype, one related to the perceiver and one related to the target. The first is the perceiver’s sociopolitical beliefs about how to approach

[1] The traits that most align with those of scientists are likely the “masculine cognitive” stereotypes discovered by Cejka and Eagly (1999). These included traits such as “good with numbers, analytical, good at problem solving, good at reasoning, quantitatively skilled”, and “mathematical.”

gender differences (gender ideology), and the second is target feminine facial appearance (gender prototypicality).

Gender Ideology

Hunt's controversial statement about the "trouble with girls", and the reactions that have ensued, are part of a lively national, international, and academic debate around how best to approach the characterization of group differences. This discussion has primarily pertained to racial and ethnic group differences (Crisp & Turner, 2011; Park & Judd, 2005; Plaut, 2010; Plaut, Thomas, & Goren, 2009; Rattan & Ambady, 2013; Verkuyten, 2005; Verkuyten, 2010), with the two most prominent perspectives being multicultural versus colorblind ideologies (Plaut, 2010). These perspectives respectively espouse recognizing and celebrating group differences versus deemphasizing group differences and treating everyone "the same" (Park & Judd, 2005; Wolsko, Park, & Judd, 2006; Wolsko, Park, Judd, & Wittenbrink, 2000). Some work in this domain has also described an assimilation perspective, which maintains that the non-dominant social group should conform to the dominant social group (Plaut et al., 2009; Verkuyten, 2005).

With the exception of a single paper, empirical research addressing *diversity science*—"how people create, interpret, and maintain group differences among individuals, as well as the psychological and societal consequences of these distinctions" (Plaut, 2010)—has focused almost exclusively on race and ethnicity. As scholars have correctly pointed out, however, many of the same issues of whether it is best to call attention to or divert it away from group differences are also relevant to discussions concerning gender (Koenig & Richeson, 2010; Plaut, 2010). This discussion is also clearly on display in a variety of popular books either celebrating (e.g., "Men are from Mars, Women are from Venus"; Gray, 1994) or lambasting gender differences (e.g., "Delusions of Gender: How our Minds, Society, and Neurosexism Create Difference; Fine, 2010). These issues track a long-standing debate among feminist scholars over how best to characterize the relationship between the genders (Crawford & Unger, 2004; Eagly, 1995; Kimball, 1994): In the quest for gender equality, is it best to argue women are no different than men, or to argue for different but equally valuable characteristics? Is it better to recognize and emphasize gender differences or to argue such differences are inflated and that we should instead be blind to them?

Drawing on existing work on ethnic ideologies, Koenig and Richeson (2010) developed a scale to measure individual differences in endorsement of what they term a sexaware versus sexblind perspective. Their 12-item scale could not differentiate between the two perspectives and so a single “sexblindness” score was computed. They argued that a “sexblind” ideology (i.e., ignoring differences) is likely to be advocated in a work setting in an effort to treat women and men equally and thereby avoid gender discrimination. Although in an absolute sense they did not find greater endorsement of a sexblind over sexaware perspective in a work context (i.e., the mean on the sexblindness scale did not differ from the scale midpoint), participants expressed greater endorsement of a sexblind perspective in a work setting compared to a social setting.

As far as I am aware, this is the only empirical work on gender ideology. This is despite a rich literature showing that in the ethnic and racial domain, emphasizing group similarities versus differences can have substantial consequences for stereotyping and bias (Correll, Park, & Smith, 2008; Gutiérrez & Unzueta, 2010; Plaut et al., 2009; Purdie-Vaughns, Steele, Davies, Ditlmann, & Crosby, 2008; Richeson & Nussbaum, 2004; Vorauer, Gagnon, & Sasaki, 2009; Wolsko et al., 2000); that there are important and meaningful differences among individuals and groups who endorse the different ideologies (Verkuyten, 2005, 2010; Wolsko et al., 2006); that ideologies can be manipulated in a manner to defend desired positions (Knowles, Lowery, Hogan, & Chow, 2009); and that there are many real-world examples of arguments and policies supporting the various ideologies (Fine, 2010, 2011; Halpern et al., 2007; Kaiser et al., 2013). For these reasons and others, it is important to explore the nature of gender ideologies.

The Four-Fold Table of Gender Ideology

Recent work in our lab has expanded diversity science in two ways: first by applying it to gender, and second by recognizing that both minimizing and emphasizing group differences can entail either good or ill will towards subordinate group members (Hahn, Banchefsky, Park, & Judd, under review; Hahn, Judd, & Park, 2010). Although the original ethnic/racial ideology manipulations developed to instantiate a multicultural or colorblind perspective (Wolsko et al., 2000) contained strong prescriptive language emphasizing the importance of intergroup harmony and equality (and therefore encouraging positive evaluations of ethnic outgroups), decades of research on prejudice and intergroup hostility indicate that intergroup relations are often contentious in nature and that intergroup harmony and equality may not be a universal goal.

The *Four-Fold Model of Intergroup Ideology* posits that four distinct intergroup ideologies arise by crossing two dimensions: 1) the extent to which people think that group distinctions (e.g., ethnicity, gender) are meaningful and important in how one ought to treat others versus the extent to which people think that individuals should be treated the same regardless of their group; and 2) whether sentiments towards the subordinate group (i.e., ethnic minorities, women) are largely positive and accepting versus negative and critical. The four distinct perspectives that make up gender ideology are depicted in Table 1. The *gender blind* ideology is defined by the desire to avoid category distinctions, and to treat people as individuals rather than as members of their gender category. Importantly, it is also defined as an ideology where others, regardless of their gender category membership, are responded to positively and respectfully. It argues, “We can all get along, if we just treat each other as valued human beings without regard to gender group membership.” A more negative instantiation of the “blind to gender” perspective is an *assimilationist* ideology. This, from the point of view of the dominant group, is a prescriptive and even hegemonic perspective arguing that, “We are all the same,” but that sameness is defined by the dominant group norms. Because social discourse has primarily focused on inclusion and acceptance of women in traditionally male-dominated domains, assimilationism was formulated as women achieving success by assimilating to male norms. Importantly, assimilationism recognizes that differences between the genders exist but advocates minimizing or deemphasizing them by having women adapt to masculine norms in male-dominated environments. Women will be treated fairly and without reference to their category membership in masculine work environments so long as they assimilate to male culture (i.e., “We can all get along so long as you become like us”).

The *gender aware* ideological viewpoint is defined by positive regard towards women, coupled with the recognition of category distinctions and a desire to preserve them to build a strong society that accommodates group differences. It argues, “We can all get along, respecting our differences, and live in a society that is stronger because of our diversity.” But this too has a more negative version: a *segregationist* ideology argues that the group differences are so large that we are better off if we simply stick to our separate spheres of expertise. It argues that, “The differences between us are so great that society will function optimally if we segregate group members,” which, given that men occupy a position of greater status and access to resources, functionally serves to maintain existing inequalities and keep women in lower status or less desirable positions. This perspective most aligns with Tim Hunt’s advocacy for single-sex labs in light of the “trouble with girls.”

		Emphasis on Group Distinctions (Differentiation)	
		Low – Distinctions should be minimized	**High** – Distinctions should be emphasized
Evaluations of Subordinate Group Members	**Positive**	Gender Blindness	Gender Awareness
	Negative	Assimilationism	Segregationism

Table 1. The four-fold Gender Ideology table.

This conceptualization of gender ideology clarifies the feminist debate about how to approach gender difference. Opponents to gender blindness claimed that this perspective could translate into devaluing feminine characteristics and pressuring women to be like men (i.e., assimilationism). On the other hand, opponents to gender awareness were concerned that this perspective could overemphasize and essentialize gender difference, justifying traditional gender roles (i.e., segregationism; Kimball, 1994). Ultimately, each perspective has a positive instantiation aimed as producing gender equality, and each has a more negative instantiation that ultimately bolsters the status quo.

We developed a psychometrically sound instrument to measure individual differences in endorsement of these four ideological perspectives (Hahn et al., under review). Testing multiple versions of the scale (as well as a parallel scale for measuring Ethnic Ideology) resulted in a final scale that has adequate internal consistency and clear separation of the four perspectives. It consists of 4-5 items to measure each ideology, for a total of 18 items (see Table 2). Participants respond to each item in a randomized order on a scale from 1 (strongly disagree) to 7 (strongly agree). A confirmatory factor analysis (CFA) confirmed that a four-factor model was superior to models with fewer factors, and that there was measurement invariance for men and women, meaning that men and women interpreted the items similarly (Cheung & Rensvold, 2002). Having developed and validated the scale, we examined meaningful correlates.

Gender Blind Items

1. You can find commonalities with every person no matter what their gender is.
2. All humans are fundamentally the same, regardless of their gender.
3. In order to achieve a harmonious society, we must stop thinking of men and women as different from each other, and instead focus on what makes us similar.
4. It is important to pay attention to the individual characteristics that make a person unique rather than his or her gender.

Gender Aware Items

1. Learning about the different ways that men and women resolve conflict will help us create a more harmonious society.
2. The differences between men and women should be acknowledged and celebrated.
3. If we want to help create a harmonious society, we must recognize that men and women have a right to maintain their own unique perspectives.
4. We must appreciate the unique characteristics of men and women in order to have a cooperative society.
5. Men and women have different but equally useful ways of accomplishing tasks.

Assimilationism Items

1. Children from both genders should be taught that success in the business world comes from adopting masculine personality qualities.
2. Women in the corporate world should embrace a masculine work ethic.
3. In order for the American workforce to be internationally competitive, women must better adapt to the ways of masculine corporate culture.
4. If a woman decides to enter a traditionally masculine field, she will be more successful if she adopts the prevailing male customs and behaviors.

Segregationism Items

1. Having men and women work side-by-side increases the likelihood of conflict.
2. Boys and girls have different learning styles and therefore it makes sense if they go to separate schools.
3. People are naturally more comfortable working and interacting with others of their same gender.
4. Men and women are naturally suited to different jobs and should continue to do those.
5. It is important to maintain some all male and all female groups to preserve gender specific interests and traditions.

Table 2. Gender Ideology items.

Correlates of Gender Ideology

Endorsement of the abstract ideologies was related to concrete beliefs about how gender-related conflicts in male-dominated environments should be resolved (Hahn et al., under review). Specifically, participants read about a female engineer who felt that her male colleagues were not taking her seriously because of her preference to dress in a feminine style of clothing and wear make-up. They then rated their agreement with four possible solutions, each written to embody one of the four gender difference ideologies (e.g., assimilationism:

Karen probably needs to realize that if she wants to be taken seriously and succeed in a field dominated by men, she might have to act the part by dressing and behaving as much as possible like her male co- workers). The rated efficacy of each solution was regressed onto endorsement of the four ideologies, participant gender, and political orientation. In all cases, the best predictor of rated efficacy was the respondent's endorsement of the matching ideological perspective.

The ideologies also predicted beliefs about the importance of social categories. In multiple regressions that again assessed each ideology's predictive value over and above the other, gender awareness and segregationism were related to viewing group differences as larger and more meaningful, whereas gender blindness was related to assigning less value to social group differences (e.g., *Humans divide the world into different categories because group differences are real and meaningful).*

Other work in our lab has shown that the gender ideologies shift depending on the gender composition of one's academic major. In a study of nearly 2000 undergraduates, the percentage of men in one's major at the University level was positively related to personal endorsement of assimilionatism and segregationism, and negatively related to endorsement of gender blindness (over and above participant gender and its interaction with the proportion of men in the major). Moreover, agreement with gender-STEM stereotypes (i.e., "I generally expect men tend to do better in math and science than women") was positively related to assimilationism and segregationism, and negatively related to gender blindness (Banchefsky & Park, in preparation).

This foundational work motivated the central question in the present research: do the gender ideologies moderate perceptions that men as more "cut out" for science than women—that is, that men possess more of the traits characteristic of scientists? In other words, are perceptions of whether the genders "fit" with a science career moderated by individual differences in gender ideology? I hypothesized that on average, men would be ascribed more scientist stereotypic traits, and fewer scientist counter-stereotypic traits, relative to women (this will be referred to as a *gender-science stereotype* to reflect that the trait stereotypes of scientists are viewed as differentially prevalent for the genders). In addition, I hypothesized that this gender-science stereotype would be stronger among those who increasingly endorse assimilationism and segregationism, and weaker among those who increasingly endorse gender blindness.

Gender Prototypicality

The second factor invested was whether or not the target's femininity, or gender prototypicality, affected the degree to which she elicited gender-science stereotypes. Although gender disparities within STEM are well-established, the present research examined how a subtle cue that varies *within* each gender might influence women's trajectory in STEM domains. Specifically, we ask whether naturalistic variation in the feminine appearance of women and men affects the perceived likelihood that the person is a scientist.

The perceived incompatibility between femininity and science is a known 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 "view that women in science or technical fields are unfeminine" is a problem for women pursuing these careers; indeed the more that women perceived that this was a problem, the less satisfied they were in their field (Hartman & Hartman, 2008). Women in STEM environments have reported feeling unable to present themselves in stereotypically feminine ways (e.g., wearing a skirt, expressing emotions) because they do not want to draw attention to their gender or are apprehensive that they will be deemed unsuitable for a STEM career (e.g., Hewlett et al., 2008; Pronin et al., 2004). Moreover, conveying that women in STEM can be overtly feminine doesn't provide a simple solution to the problem. A STEM female role model who performed feminine-stereotyped behaviors (e.g., wore pink clothing, liked reading fashion magazines) was uninspiring to adolescent women because being both good at math and science and simultaneously feminine seemed unattainable (Betz & Sekaquaptewa, 2012).

In order to blend in with their male colleagues and avoid being judged as unsuited for STEM, women in STEM have reported assimilating—minimizing their feminine appearance, eschewing feminine traits and goals (e.g., desire to have children, Pronin, Steele, & Ross, 2004), and distancing from other feminine women (Rhoton, 2011), practices that are isolating and can lead to abandoning field altogether (Hewlett et al., 2008). In light of these accounts, I examined whether women with feminine facial appearance and men with masculine facial appearance elicited stronger implicit gender-science stereotypes than their low prototypic counterparts (women with more masculine and men with feminine facial appearance). Indeed, facial appearance is one cue that women carry with them that may signal degree of fit to STEM.

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 & Todorov, 2006). Like the major social categories of gender and race (Ito & Urland, 2003), variations in facial appearance are automatically processed and automatically activate stereotypes: stronger stereotypic Black features (e.g., broad nose, thick lips) activate Black stereotypes (e.g., close with family, failing school; Blair, Judd, & Fallman, 2004), and baby-faced features (e.g., large round eyes and foreheads) activate youthful characteristics (e.g., naïve, submissive; Zebrowitz, Tenenbaum, & Goldstein, 1991). Evidence also suggests that gender prototypical appearance has important consequences. For example, it is the strongest activator of other gender stereotype components—traits, role behaviors, and occupations. That is, targets described has having a feminine physical appearance were assumed to possess feminine traits, roles, and occupations as well (Deaux & Lewis, 1984). Moreover, Sczesny and Kühnen (2004) showed that males and females with feminine appearance were judged as less suitable for leadership positions than their masculine counterparts (see also Friedman & Zebrowitz, 1992; Zebrowitz et al., 1991).

Implicit evaluations have also been shown to be sensitive to facial features (Blair, Judd, & Chapleau, 2004; Blair, Judd, & Fallman, 2004; Blair, Judd, Sadler, & Jenkins, 2002; Deaux & Lewis, 1984; Maddox, 2004). Livingston and Brewer (2002) found in a sequential priming procedure that more Afrocentric Black individuals elicited greater negativity than their less Afrocentric counterparts (Livingston & Brewer, 2002). Using the same task, Ito and colleagues (2011) demonstrated that implicit negativity was sensitive to race in a linear fashion, with the greatest negativity elicited by Blacks, followed by racially ambiguous targets, followed by Whites (Ito, Willadsen-Jensen, Kaye, & Park, 2011). Notably, Livingston and Brewer (2002) did not find any evidence that automatic *stereotypes* of Blacks were affected by racially prototypical features. However, other implicit tasks have shown that implicit stereotypes are sensitive to the specific stimuli; Macrae et al. (2002) used sequential priming to show that gender stereotypes (regarding gender stereotypic attributes such as Jeep and lingerie) were weaker with unfamiliar (e.g., Isaac, Glenda) than with familiar names (e.g., John, Sarah; Macrae, Mitchell, & Pendry, 2002). Finally, well-liked Black exemplars (contrasted with disliked Whites) elicited less negativity than disliked Black exemplars (contrasted with well-liked Whites) in an IAT (Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003).

Feature-based stereotyping, or stereotyping based on within-category features, is important because people are not as aware of it, making it less amenable to conscious correction or control (Blair et al., 2004b; Sceszny & Kühnen, 2004). Feature-based stereotyping can affect judgments even when social categories do not. In research by Sczesny and Kühnen (2004), men were rated as better leaders than women only when participants were under cognitive load, suggesting that participants who were not under cognitive load adjusted their responses to avoid gender bias. In contrast, a bias against feminine appearance was present regardless of cognitive load, suggesting that it was operating without participants' awareness and/or beyond their control.

Similarly, Blair and colleagues found Black and Whites who possessed more Afrocentric facial features (e.g., darker skin, wider nose, thicker lips) activated stronger Black stereotypes in an automatic fashion (i.e., participants were unaware of the bias and unable to control it; it also occurred efficiently, even when cognitive resources were constrained by having participants do another task simultaneously; Blair et al., 2004b). Such automatic stereotyping can have severe consequences—Black and White inmates who were more Afrocentric received harsher sentences (controlling for severity of the crime) despite the lack of a race category bias in criminal sentences. Although judged were likely aware of and sensitive about expressing categorical racial bias in their sentences, they were not aware or able to control being swayed by Afrocentricity (Blair et al., 2004a; see also Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Kahn & Davies, 2011).

In summary, research has shown that gender stereotypes contribute to the gender gap in science, technology, engineering, and math domains (STEM), but this work has not investigated whether men and women are perceived as differentially possessing *traits* stereotypic of scientists (i.e., gender-science stereotypes), nor whether such stereotypes depend on the perceiver's sociopolitical beliefs about how to approach and deal with gender differences (i.e., gender ideologies) or the target's gender prototypical appearance. The present research aimed to address these gaps in the literature.

CHAPTER 2: Study 1

The primary purpose of Study 1 was to pre-test photos of real male and female scientists who were high on gender prototypicality (i.e., feminine women and masculine men) and low on gender prototypicality (i.e., masculine women and feminine men) to use as stimuli in the lab portion of the study. A secondary goal was to examine the impact of facial femininity of perceived likelihood of being a scientist. Thus in addition to collecting femininity ratings of each face, participants judged the likelihood that each person was a scientist as well as an early childhood educator, allowing an examination of the relationship between judged femininity and judged likelihood of being a scientist. The hypothesis was that female scientists with feminine facial appearance would be judged as less likely to be scientists by naïve participants. There was not a clear hypothesis for men; feminine men may activate feminine gender stereotypes as well, but they also trigger "nerdy" male stereotypes (Cheng, 2008) that align with stereotypes about the types of people who populate STEM domains (Cheryan, Siy, Vichayapai, Drury, & Kim, 2011), thereby perhaps increasing their judged likelihood of being scientists.

To rigorously test this hypothesis, we attended to the conceptual and methodological issues of stimulus sampling (Wells & Windschitl, 1999) by employing a large stimulus set and by treating the stimuli as random in the analysis (Judd, Westfall, & Kenny, 2012). Relative to prior research on the effects of facial femininity that typically employs only a small sample of stimuli (e.g., six photographs, Sczesny & Kühnen, 2004), the present approach offers several advantages. Specifically, we:

(a) Treated variation in masculinity-femininity continuously, sampling from and analyzing the full spectrum rather than dichotomizing this into simply feminine versus masculine faces;

(b) Used a large sample of faces of real people, and specifically, tenured or tenure-track faculty members in STEM departments at elite U.S. universities;

(c) Treated both faces (stimuli) and participants as random factors (Baayen, Davidson, & Bates, 2008; Judd et al., 2012), in contrast to all of the existing work we were able to identify in which faces were treated as a fixed factor. This analysis permits generalization from this specific sample of faces to other samples of faces that might have been used (Clark, 1973; Judd et al., 2012).

In the first study reported, every participant evaluated every stimulus *both* on the outcome variable (likelihood of being a scientist) as well as perceived masculine-feminine appearance. As described below, this permits parsing the variation in both the predictor and response variables into separate components examining (a) whether faces that are judged as more feminine on average are judged as less likely to be scientists, (b) whether perceivers who judge all faces as more feminine on average perceive all faces as less likely to be scientists, and (c) when a particular perceiver judges a particular face as even more feminine than it is judged on average, and than the perceiver judges faces on average, does he or she perceive that face as less likely to be a scientist? In other words, we can examine the influence of facial femininity on judged career likelihood while statistically holding constant both the perceiver and the target.

Study 1a

Participants

Participants were 55 United States based workers on Amazon.com's Mechanical Turk (25 men, 26 women; 78% White, 13.72% Asian, 7.8% Black, 3.9% Latino; *Mean* age =33.78, *SD*=13.16) who were compensated $0.50 for their time. 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 removed.

Stimuli

Eighty photographs (40 men, 40 women) of tenured or tenure-track faculty in male-dominated STEM disciplines, working in elite STEM departments in United States' universities, served as the study stimuli. Programs were selected according to U.S. News and World Report's rankings of the best STEM graduate programs. In total, eleven research universities (e.g., MIT, Caltech, Harvard, UC Berkeley), and fifteen STEM programs (e.g., computer science, physics, engineering) were represented. High quality, color photos were selected from the program websites; each included the individual's entire face and shoulders. In order to avoid variation in responses due to race, all faculty selected were White. All were also making direct eye contact and smiling (see Figures 1 and 2 for example stimuli). Because our critical question concerned variation in gender prototypicality, we intentionally selected faces to maximize variation on the spectrum from masculine to feminine by attending to such factors as facial structure and hair, hairstyle and length, presence of make-up, and jewelry.

Figure 1. Example stimuli of women used in Study 1. Feminine women (around +1 SD, top rows), average women (around the mean, middle row), and masculine women (around -1 SD, bottom row).

Figure 2. Example stimuli of men used in Study 1. Feminine men (around +1 SD, top rows), average men (around the mean, middle row), and masculine men (around -1 SD, bottom row).

Procedure

As a study on “first impressions,” participants were told that first impressions are made very quickly and are often surprisingly accurate (e.g., “In one study, people were able to detect someone’s sexual orientation by looking at their face for just half a second (Rule, Ambady, Adams, & McCrae, 2008).”). Participants were told to go with their ‘gut reactions’ and asked to rate each photo on three 7-point scales: *masculine* to *feminine*, *likable* to *unlikable*, and *unattractive* to *attractive*, always in this order (see Appendix A for all Study 1 materials). Next, they estimated the likelihood that the individual was a scientist, followed by the likelihood that the person was an early childhood educator (ECE, a stereotypically female profession that is highly female-dominated, 97% women; Carnevale, Strohl, & Melton, 2013), on 6-point scales (*very unlikely* to *very likely)*. Finally, they estimated the age of the target, selecting one of eight 5-year ranges starting at 25 and ending at 60 and above*.* Target gender of the photos was blocked and counterbalanced. Photos within each block were randomized, and each was presented on a separate screen. Participants were not aware that the photographs were all of actual STEM academics in the United States. Lastly, participants completed demographics, which included their gender, age, ethnicity, educational attainment, political orientation, and career.

Data Analysis

Initial analyses examined whether participant gender or target gender order (i.e., whether participants rated men or women first) affected any of the judgments. No effects were detected, so analyses collapsed across these variables. To examine whether femininity impacted career ratings, career likelihood was analyzed as a function of career type (science vs. educator, contrast coded), face gender (male vs. female, contrast coded), femininity (mean centered), and all possible interactions. Data were analyzed using linear mixed models with crossed random effects of participants and stimuli (Baayen, Davidson, & Bates, 2008; Judd et al., 2012). These mixed models contained all possible random intercepts and slopes (Barr, Levy, Scheepers, & Tily, 2013); the covariances between the random effects were not estimated due to convergence problems with the full model that included the covariances (which entailed estimating an additional 81 parameters).

Because our experimental design involved every participant providing continuous ratings of femininity and career likelihood for every face, the femininity predictor varied both between and within participants as well as between and within faces. It was therefore possible to decompose the influence of the femininity predictor into three distinct effects based on the different sources of the predictor variance. These three effects will be referred to as the *target effect*, the *perceiver effect*, and the *relationship effect* of femininity, 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 affects judgments of career likelihood. It is computed as $\overline{F_{\cdot j}}$ 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 affects judgments of career likelihood. It is computed as $\overline{F_{i\cdot}}$ and then mean centered in the mixed model.

Case 3. *Relationship* (i.e., Target × Perceiver interaction): The relationship effect 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 $F_{ij} - \overline{F_{\cdot j}} - \overline{F_{i\cdot}}$ and then mean centered in the mixed model (see Raudenbush, 2009; Rosnow & Rosenthal, 1991).

Including each of these predictors allows an appraisal of how stimuli, perceivers, and the relationship between them each uniquely contribute to 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 & Gelman, 2006; Bell & Jones, 2015).

We predicted that the three-way interaction of interest (Target Gender × Career × 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 ECE (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 ECE (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 the problem of biased parameter estimates mentioned above.

Results

Preliminary Analyses.

In preliminary analyses that treated face as the unit of analysis and averaged across participants, we explored whether the stimulus set for men and women were perceived equivalently. Both groups were perceived as about the same age, equally likable, and equally attractive (Table 3). Unsurprisingly, women were rated as significantly more feminine than men and also significantly more likely to be an early childhood educator. Notably, women were rated as equally likely as men to be scientists, perhaps reflecting our stimuli selection in which the genders were in fact equally likely to be scientists (Table 3).

	Mean (SD)	
Feature Dimension	Target Women	Target Men
Masculine - Feminine	5.11 (1.03)	2.85 (.78)[a]
Unattractive - Attractive	4.25 (.92)	4.04 (.63)
Unlikable - Likeable	3.94 (.40)	3.92 (.44)
Age	3.11 (1.16)	2.80 (1.16)
Likelihood Scientist	3.90 (.43)	3.96 (.53)
Likelihood Early Childhood Educator	3.93 (.37)	3.14 (.34)[a]

Table 3. Mean ratings by target gender on six dimensions in Study 1a. Scales were from 1-7 for the first three dimensions (masculine - feminine, likeable – unlikable (reverse scored above), and unattractive - attractive); and from 1 (very unlikely) to 6 (very likely) for career ratings. Age ratings were categorical and represented ranges (1 = 25-29 years; 2 = 30-35 years; 3 = 36-40 years; 4 = 41 to 45 years; 5 = 46 to 50 years; 6 = 51-55 years; 7 =56-60 years; and 8 =61+ years). [a] Indicates that male faces were different from female faces at $p < .01$ level.

As depicted in Table 4, average judged femininity and attractiveness were highly positively correlated for women and less strongly but negatively correlated for men, replicating previous research (Perrett et al., 1998). Femininity was also negatively correlated with age, but more so for men than for women. Due to the extremely high correlation between femininity and attractiveness for women, and collinearity problems this creates in the predictors, we did not include both simultaneously in analyses, instead analyzing their effects in separate models. Additional ancillary analyses examined age and femininity simultaneously.

Feature Dimension	Female Faces 2.	3.	4.	5.	6.	Male Faces 2.	3.	4.	5.	6.
1. Feminine	.89**	.59**	-.41**	-.56**	.75**	-.48**	.28	-.69**	-.13	-.01
2. Attractive	--	.69**	-.63**	-.61**	.67**	--	.68**	-.08	-.29	.54**
3. Likeable		--	-.45**	-.25	.65**		--	-.11	.11	.76**
4. Age			--	.31*	-.31*			--	.50**	-.37*
5. Scientist				--	-.65**				--	-.25
6. ECE					--					--

Table 4. Correlations among face judgments by target gender in Study 1a. Correlations are based on averages for each face (N = 40 for men, N = 40 for women). Age ratings were categorical and represented age ranges (1 = 25-29 years; 2 = 30-35 years; 3 = 36-40 years; 4 = 41 to 45 years; 5 = 46 to 50 years; 6 = 51-55 years; 7 = 56-60 years; and 8 = 61+ years).*$p < .05$; **$p < .01$.

Femininity and Career Likelihood.

Table 5 presents fixed effects output, including the critical, hypothesized interactions. First, the hypothesized three-way interaction between Target Gender, Career Domain, and Femininity emerged (target effect), $F(1, 97.9) = 6.10$, $p = .015$.[2] We broke down the interaction by target gender, examining women and men separately. Whereas perceived femininity affected career judgments for women (Career Domain × Femininity, $F(1, 90.7) = 25.00$, $p < .001$), it had no impact on career judgments for men (i.e., Career Domain × Femininity was not significant for men, $p = .53$). This is shown by the bold-type regression lines in each panel of Figure 3, where the slopes in the two panels for women (but not men) are different from each other, and both significantly different from zero. In line with the hypothesis, as the average rated femininity of a woman increased, she was judged as significantly less likely to be a scientist, $F(1, 78.3) = 12.67$, $p < .001$, and significantly more likely to be an early childhood educator, $F(1, 74.3) = 41.99$, $p < .001$.

[2] Decimal degrees of freedom are typical in mixed model analyses that treat stimuli as random.

Effect	Estimate	*SE*	df	*t*	df	*p*
Intercept	3.69	81.3	.075	49.09	81.3	<.001
Main Effects						
Career	.33	112	.070	4.69	112	<.001
Gender	.192	95.3	.046	4.18	95.3	<.001
Feminine_Face (Target effect)	-.01	78.4	.030	-.37	78.4	.715
Feminine_Ss (Perceiver Effect)	.144	1.25	.116	47.6	1.25	.218
Feminine_Rel (Relationship Effect)	.032	43.5	.013	2.38	43.5	.022
Two-Way Interactions						
Career × Gender	-.053	95.3	.061	-.87	95.3	.385
Career × Feminine_Face (Target effect)	-.145	93.3	.042	-3.41	93.3	<.001
Career × Feminine_Ss (Perceiver Effect)	-.121	-1.49	.081	45.7	-1.49	.142
Career × Feminine_Rel (Relationship Effect)	-.047	59.6	.017	-2.83	59.6	.006
Gender × Feminine_Face (Target effect)	.030	85.8	.032	.96	85.8	.339
Gender × Feminine_Ss (Perceiver Effect)	-.049	46.6	.040	-1.22	46.6	.228
Gender × Feminine_Rel (Relationship Effect)	.030	51.7	.013	2.27	51.7	.028
Three-Way Interactions						
***Career × Gender × Feminine_Face (Target effect)**	**-.108**	**97.9**	**.044**	**-2.47**	**97.9**	**.015**
Career × Gender × Feminine_Ss (Perceiver Effect)	.064	40	.052	1.24	40	.221
***Career × Gender × Feminine_Rel (Relationship Effect)**	-.042	**48.5**	**.021**	**-1.99**	**48.5**	**.052**

Table 5. Mixed-models results for fixed effects for career likelihood judgments in Study 1a. SE =standard error. df = Satterthwaite approximate degrees of freedom. *Indicates hypothesized effect of interest.

A three-way interaction also emerged between Target Gender, Career Domain, and Femininity (relationship effect), $F(1, 48.5) = 3.96$, $p = .052$. This is shown by the thin, shorter regression lines in each panel of Figure 3, which depict the within-stimulus regressions of career likelihood on femininity (with perceiver effects removed) for each face.

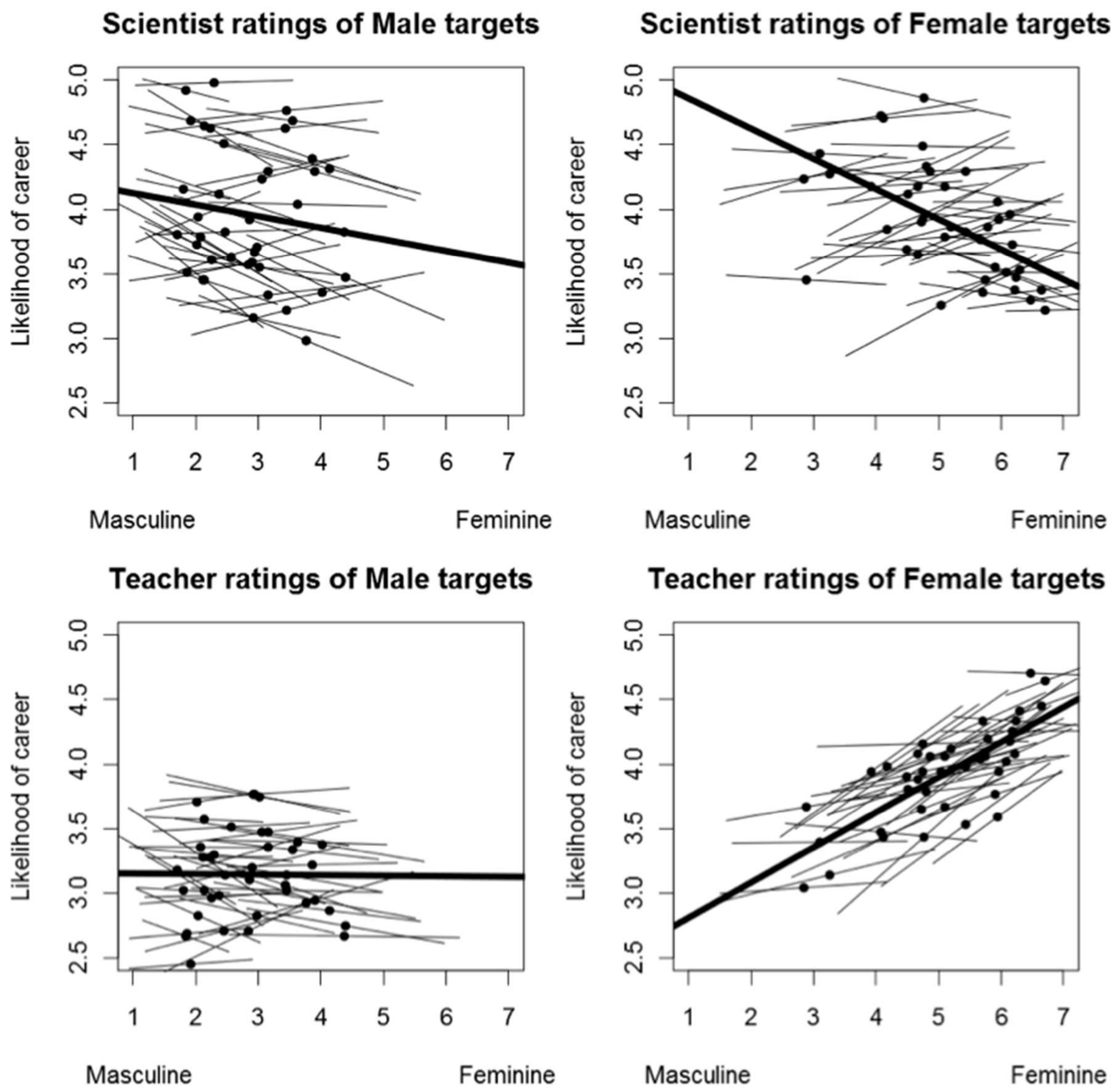


Figure 3. Plot of the relationship between femininity and career likelihood by target gender and career (Study 1a). The points in each panel represent the mean femininity and likelihood ratings for each stimulus face (i.e., the target effects), and the bold regression line in each panel of career likelihood 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.

As before, the relationship effect of femininity affected career judgments for women, $F(1, 66.2) = 12.60$, $p < .001$, and had no impact on career judgments for men (i.e., Career Domain × Femininity was not significant for men, $p = .80$). Breaking this down further by career indicated that the interaction for women was driven by early childhood educator ratings: when a given participant viewed a given woman as more feminine than expected (based on the perceiver's typical femininity rating and the face's typical femininity rating), he or she also rated that woman as more likely to be an early childhood educator, $F(1, 5918) = 38.81$, $p < .001$. However, perceiving a given face as more feminine than expected (for that participant and for that face) did not affect the perceived likelihood of being a scientist, $p = .97$. Thus the thin regression lines in Figure 3 show a significant positive slope on average only for the early childhood educator ratings of the female targets. There were no significant femininity perceiver effects. Lower order effects emerged in the model, but all were qualified by the two reported three-way interactions (see Table 5).

Ancillary analyses.

We also examined a model that controlled for the target's perceived age, again parsed into three sources of variation (perceiver, stimuli, and the interaction) and centered appropriately. There were no significant effects of age, and importantly, the critical three-way interactions involving target gender, career type, and femininity remained significant even when controlling for age.

As discussed above, target mean attractiveness and femininity were very highly correlated (.89) for female targets. Because of potential collinearity problems, models that included both as simultaneous predictors were avoided. Instead we estimated a model parallel to that estimated for femininity, in order to examine whether the attractiveness effects were the same as or different from those due to femininity. These models revealed significant two-way interactions between Career likelihood × Attractiveness (again for both target effect and relationship effect) indicating that more attractive targets were seen as less likely to be scientists and more likely to be ECE (target effect: $F(1, 105) = 28.09$, $p < .001$; relationship effect: $F(1, 55.7) = 10.86$, $p = .002$). These effects did not depend on target gender, and align with stereotypes of scientists as unattractive (Hannover & Kessels, 2004). The fact that these effects did not depend on target gender suggests that they are very different in form from the femininity effects.

Study 1b

One concern in Study 1a is that asking participants to evaluate the targets' appearance (e.g., attractiveness, femininity) may have made these concepts especially salient in making the career judgments. Would the effects emerge even when these were less salient to perceivers? Another potential concern is that the blocked presentation of the stimuli by gender lacked external validity and produced excessive attention to within-category variations in appearance. Finally, perhaps rating just two careers forced participants to make a trade-off in judging the targets. Study 2 addressed these concerns in addition to testing the reliability of the effects.

Participants

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; M age = 36.27, SD = 11.41) in order to account for the addition of a between subjects condition. The survey took about 20-30 minutes to complete and workers were paid $.75.

Design

Study 1b replicated Study 1a with the following differences: 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 female faces followed by all male faces or vice versa, with the order of the target gender blocks counterbalanced; faces were presented in a randomized order within each block. In the mixed condition, all faces were presented in a fully randomized for each participant. Second, participants only made career likelihood judgments of each target, ensuring that initial explicit judgments of appearance would not influence career likelihood judgments. Third, to make it less apparent that we were examining a male stereotypic (scientist) and female stereotypic career (teaching), participants first rated a relatively gender-neutral career, journalist (64% female; Carnevale et al., 2013), for each target.

Procedure

The cover story was very similar to Study 1a (see Appendix A). 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, in that order, and again on 6-point scales (*very unlikely* to *very likely*). Participants lastly completed the same demographics as in Study 1a.

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), face gender (men vs. women, contrast coded), femininity (mean centered), presentation (blocked vs. mixed, contrast coded), participant gender (men vs. women, contrast coded) and all possible interactions. For the femininity predictor, the average femininity rating of each face from Study 1a was employed. We predicted that the three-way interaction of interest (Target Gender × Career × Femininity) 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.

Data were again analyzed using linear mixed models with crossed random effects of participants and stimuli, which contained all possible random intercepts and slopes, but not covariances. Because perceivers in this study did not make femininity ratings, we could not decompose femininity into effects due to target, perceiver, and their interaction as in Study 1a; instead, all femininity effects reported below are “target effects.”

Results

Femininity and Career Likelihood.

Table 6 presents the fixed effects of our hypothesized interactions. First, the predicted three-way interaction between Target Gender, Career Domain (Science vs. Other), and Femininity again emerged, $F(1, 81.4) = 26.94$, $p < .001$. Whereas perceived femininity affected career judgments for women (Science Vs. Non-Scientist × Femininity, $F(1, 78.5) = 43.16$, $p < .001$), it again had no impact on career judgments for men (i.e., this two-way interaction was not significant for men, $p = .11$, see Figure 4).

Once again supporting the hypothesis, as the average rated femininity of a woman increased, she was judged as significantly less likely to be a scientist, $F(1, 76.8) = 26.83$, $p<.001$, and significantly more likely to be a non-scientist, $F(1, 77.7) = 58.37$, $p<.001$. The lack of a Teacher vs. Journalist × Target Gender × Femininity interaction indicated that femininity affected career ratings of journalist the same way that it affected ratings of early childhood educator.

Effect	Estimate	*SE*	df	*t*	*p*
Intercept	3.53	0.05	229	68.00	<.001
Main Effects (Predictors of Interest)					
Scientist v. Non-Scientist (Scientist)	0.28	0.05	100	5.23	<.001
Educator vs. Journalist (ECE v. J)	-0.11	0.06	115	-1.91	0.06
TGender	0.10	0.04	91	2.70	0.01
Femininity	0.06	0.03	78	2.54	0.01
Two-Way Interactions (Predictors of Interest)					
Scientist × TGender	-0.01	0.05	81	-0.19	0.85
Scientist × Femininity	-0.09	0.04	78	-2.68	0.01
ECE v. J × TGender	0.23	0.05	92	4.38	<.001
ECE v. J × Femininity	0.04	0.04	83	1.05	0.29
TGender × Femininity	0.17	0.03	83	6.49	<.001
Three-Way Interactions (Predictors of Interest)					
***Scientist × TGender × Femininity**	**-0.18**	**0.04**	**81**	**-5.19**	**<.001**
ECE v. J × TGender × Femininity	-0.05	0.04	86	-1.42	0.16

Table 6. Mixed-models results for predictors of interest fixed effects for career likelihood judgments in Study 1b. SE = standard error. df = Satterthwaite approximate degrees of freedom. *Indicates hypothesized effect of interest.

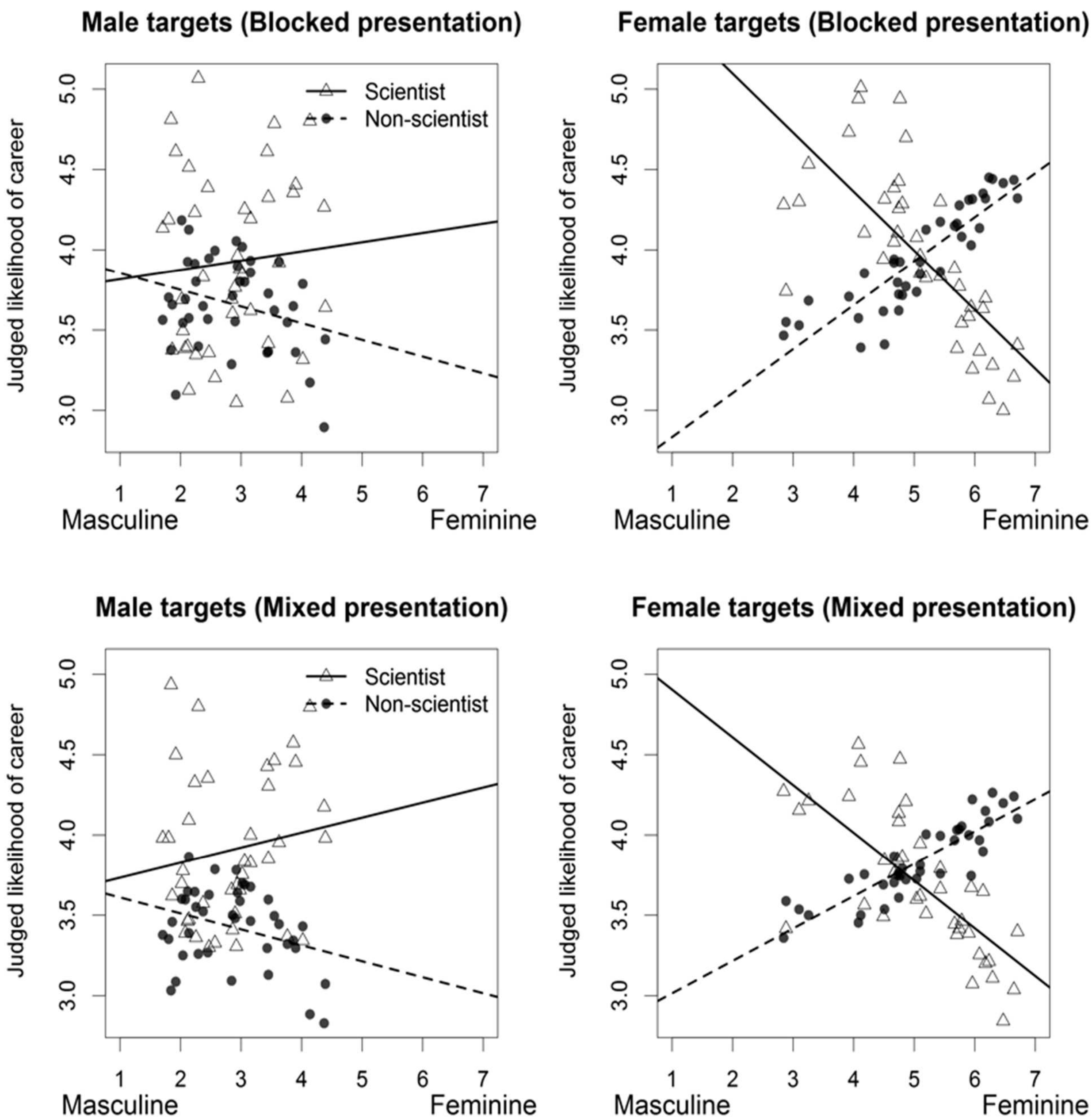


Figure 4. Plot of relationship between femininity and career likelihood by Target Gender, Career, and Presentation Type in Study 1b.

Table 7 presents fixed effects related to stimuli presentation; the femininity effect for women was marginally stronger in the blocked condition than in the mixed condition, as indicated by a marginal 4-way interaction, $F(1, 201) = 2.96$, $p = .09$.

Presentation Effects (Mixed v. Blocked)	Estimate	*SE*	df	*t*	*p*
Presentation	-0.07	0.04	222	-1.72	0.09
Presentation × Scientist	0.00	0.02	245	-0.16	0.88
Presentation × ECE v. J	-0.04	0.03	224	-1.40	0.16
Presentation × Target Gender	0.05	0.02	380	3.52	0.00
Presentation × Femininity	-0.02	0.01	511	-2.08	0.04
Presentation × PGender	0.00	0.04	222	0.09	0.93
Presentation × Scientist × Target Gender	**-0.07**	**0.01**	**226**	**-4.72**	**<.001**
Presentation × Scientist × Femininity	0.02	0.01	167	2.32	0.02
Presentation × Scientist × PGender	-0.02	0.02	230	-0.78	0.43
Presentation × ECE v. J × Target Gender	**0.07**	**0.02**	**207**	**3.39**	**0.00**
Presentation × ECE v. J × Femininity	0.00	0.01	145	-0.37	0.71
Presentation × ECE v. J × PGender	0.05	0.03	217	1.74	0.08
Presentation × Target Gender × Femininity	-0.02	0.01	279	-2.10	0.04
Presentation × Target Gender × PGender	-0.02	0.02	380	-1.27	0.21
Presentation × Femininity × PGender	-0.02	0.02	380	-1.27	0.21
***Presentation × Scientist × Target Gender × Femininity**	**0.02**	**0.01**	**201**	**1.71**	**0.09**
Presentation × Scientist × Target Gender × PGender	-0.01	0.01	241	-0.99	0.32
Presentation × Scientist × Femininity × PGender	0.01	0.01	188	1.16	0.25
Presentation × ECE v. J × Target Gender × Femininity	-0.01	0.01	167	-0.55	0.58
Presentation × ECE v. J × Target Gender × PGender	0.01	0.02	204	0.60	0.55
Presentation × ECE v. J × Femininity × PGender	-0.01	0.01	145	-0.56	0.58
Presentation × Target Gender × Femininity × PGender	0.00	0.01	279	0.03	0.97
Presentation × Scientist × Target Gender × Femininity × PGender	-0.01	0.01	145	-0.56	0.58
Presentation × ECE v. J × Target Gender × Femininity × PGender	0.00	0.01	164	0.36	0.72

Table 7. Fixed effects results for presentation (Blocked by Target Gender versus Unblocked) on Career Likelihood Judgments in Study 1b. SE = standard error. df = Satterthwaite approximate degrees of freedom. *Highest order effects of interest. Other effects of interest are in bold.

Looking within each presentation type, the simple three-way was stronger within the blocked condition, $F(1, 90.4) = 28.94$, $p < .001$, than in the mixed condition, $F(1, 89.7) = 22.09$, $p < .001$, suggesting that people may have paid more attention to women's femininity or weighed it more heavily when faces were presented in a blocked rather than mixed fashion. However, and importantly, it was significant in both presentation conditions. The femininity effect was also moderated by participant gender, $F(1, 201) = 6.15$, $p = .01$. Although the critical three-way interaction was highly significant within both genders, it was even stronger among women, $F(1, 86.4) = 32.49$, $p < .001$, than among men, $F(1, 94.8) = 19.54$, $p < .001$. This suggests that women especially may consider a woman's facial femininity as a meaningful cue of her career. However participant gender did not moderate the results in Study 1a, and so this finding should be interpreted with caution.

A variety of un-hypothesized lower order effects emerged, all of which can be viewed throughout Tables 6-8. These were all importantly moderated, however, by the predicted three-way interaction. Interestingly, the mixed presentation of men and women enhanced categorical gender biases in career judgments. That is, a significant Target Gender × Presentation × Science vs. Other interaction indicated that a categorical gender bias (that is the difference in the mean ratings for women versus men) was stronger in the mixed condition than in the blocked condition, $F(1, 226) = 22.28$, $p < .001$ (see Table 7). Simple effects indicated that the Presentation × Science vs. Other interaction was significant for both women targets, $F(1, 271) = 8.58$, $p<.01$, and men targets, $F(1, 287) = 6.25$, $p = .01$. Looking within career, women were viewed as *less* likely to be scientists in the mixed compared to the blocked presentation, $F(1, 266) = 11.56$, $p < .001$, whereas men were seen as less likely to be non-scientists in the mixed compared to the blocked presentation, $F(1, 279) = 8.64$, $p<.01$. Thus, in the mixed condition, categorical gender biases were somewhat stronger than in the blocked condition; women were seen as less likely to be scientists, and men were seen as less likely to be non-scientists.

Effect	Estimate	*SE*	df	*t*	*p*
Participant Gender	-0.01	0.04	222	-0.28	0.78
Participant Gender × Scientist	0.04	0.02	234	1.72	0.09
Participant Gender × ECE v. J	-0.02	0.03	223	-0.89	0.37
Participant Gender × Target Gender	0.00	0.02	380	0.00	1.00
Participant Gender × Femininity	0.02	0.01	511	2.16	0.03
Participant Gender × Scientist × Target Gender	0.03	0.01	234	2.26	0.02
Participant Gender × Science × Femininity	-0.02	0.01	178	-2.84	0.01
Participant Gender × ECE v. J × Target Gender	0.00	0.02	204	0.17	0.87
Participant Gender × ECE v. J × Femininity	0.02	0.01	140	2.19	0.03
Participant Gender × Target Gender × Femininity	0.01	0.01	279	1.49	0.14
***Participant Gender × Scientist × Target Gender × Femininity**	**-0.02**	**0.01**	**201**	**-2.48**	**0.01**
Participant Gender × ECE v. J × Target Gender × Femininity	0.01	0.01	163	1.07	0.29

Table 8. Remaining fixed effects results for Participant Gender on Career Likelihood Judgments in Study 1b. SE =standard error. df = Satterthwaite approximate degrees of freedom. *Highest order effect of interest.

Ancillary analyses.

When age was included as a predictor (using average age ratings for each face from Study 1), the three-way interaction remained highly significant. We also examined a model that included perceived attractiveness rather than femininity (again using average attractiveness ratings for each face from Study 1); a Career likelihood × Attractiveness interaction indicated that again, more attractive men and women were seen as less likely to be scientists and more likely to be non-scientists, $F(1, 99.6) = 94.28$, $p < .001$.

Study 1 Discussion

Two studies examined how variation in facial femininity among 80 actual scientists related to judgments about their likelihood of being a scientist. Even when physical appearance was not explicitly salient to participants (Study 1b), women scientists (but not men scientists) with more feminine appearance were seen as significantly less likely to be scientists and more likely to be non-scientists (that is, ECE or journalists). Interestingly, attractiveness was used as a career cue regardless of target gender, such that more attractive people were seen as less likely to be scientists and more likely to be non-scientists. On the other hand, perceivers only used gender typical appearance as a cue to women's, but not men's, careers. This suggests that compared to men, women may encounter additional hurdles both in pursuing and flourishing within science simply due to looking more gender typical. Before choosing science, feminine girls and women—because they don't "look" like scientists—might be treated differently by parents, teachers, and others (Tenenbaum & Leaper, 2003). These interactions may elicit a cascade of inferences that guide not only the perceiver's behavior, but that in turn affect the self-perceptions and behavior of the girls and women themselves (Snyder, Tanke, & Berscheid, 1977).

Once in a science field, women may feel that in order to be accepted by their peers, they have to suppress a basic part of their gender identity: their femininity. This may contribute to women in STEM minimizing their feminine appearance, eschewing feminine traits and goals (Pronin et al., 2004), and distancing from or criticizing other feminine women (Rhoton, 2011), practices that are isolating and can lead to abandoning field altogether (Hewlett et al., 2008). This suggests an important question that remains to be empirically addressed—because of self-selection and presentation, might women in STEM be objectively less feminine in appearance (see Carpinella & Johnson, 2013)? The present research deals with this possibility by presenting faces of tenured or tenure track faculty in a "hard science" department at top rated universities, ensuring that participants were not accurately responding to some information present in the faces that indicated poor fit for a career as a scientist.

The results were especially compelling given that variation in feminine appearance among the presented scientists was naturalistic and fairly subtle; certainly much more extreme, highly feminized portrayals of women occur throughout the popular media. That said, future research should explore whether and how various aspects of femininity (e.g., inherent facial structure or facial features vs. performed femininity such as hairstyle and make-up) might differentially contribute to inferences, attributions, and career judgments. Along these lines, it would be interesting to investigate what specific aspects of appearance contribute to overall perceived femininity.

The robust correlation between judged 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 of and less capable of controlling biases based on within-category variations in facial appearance, even when they are clearly informed about how such biases operate and asked not to use them (Blair et al., 2004b; Sczesny & 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 suggests that even in the face of clear diagnostic information that should override irrelevant cues, within-category cues nevertheless affect judgments (Blair, Chapleau, & Judd, 2005). This research suggests that an important next step would be to a) address whether variations in femininity convey information beyond likelihood of being a scientist, such as scientific competence, and b) to examine whether femininity affects judgments automatically (e.g., efficiently, without awareness; Blair et al., 2004b).

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 & Jones, 2015; Kievit, Frankenhuis, Waldorp, & Borsboom, 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 the 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—and if we were to conduct the study again with a different stimulus set of top scientists, statistically the results should replicate (Judd et al., 2012).

Finally, these findings also have implications for young women and girls who are forming academic and vocational interests. People are drawn to fields where they feel they would belong and be similar to others. In addition to being discouraged by male-dominated STEM environments (Murphy, Steele, & Gross, 2007) or those populated by male stereotypic objects (Cheryan, Plaut, Davies, & Steele, 2009), women's interest in STEM may also be thwarted by the undue perception that women scientists simply cannot be feminine. Overall, these data 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 and welcoming STEM workforce.

In sum, Study 1 provided strong evidence that subtle variations in feminine facial appearance can indeed impact judgments of likelihood of being a scientist. In Study 2, I explore whether these same scientists judged in Study 1 activate different implicit gender-science stereotypes as a function of gender, femininity, and their interaction.

CHAPTER 3: Study 2 Overview

Study 2 asks three primary questions: 1) What is the content of scientist stereotypes and are these differentially ascribed to men and women?; 2) How are implicit and explicit gender-science stereotypes moderated by gender ideologies?; and 3) How are implicit gender-science stereotypes affected by within-category variations feminine facial appearance (i.e., prototypicality)?

To address these questions, I:

1. Identified the trait characteristics that constitute the stereotype of scientists (Scientist Trait Pre-test).
2. Measured gender-science stereotypes, or the extent to which these stereotypic scientist traits were (a) explicitly ascribed to men more than to women and (b) implicitly associated with men more than with women.
3. Examined whether implicit and explicit gender-science stereotypes were moderated by individual differences in gender ideology.
4. Examined whether implicit gender-science stereotypes were moderated by gender prototypical facial appearance.

Hypotheses were as follows:

1. *Science Stereotypes:* Scientists will be stereotyped as possessing more *scientific* and *cold* traits than early childhood educators (ECE), who will be stereotyped as possessing more *unscientific* and *warm* traits than scientists.
2. *Gender-Science Stereotypes:* People will associate science stereotypic traits more with men than with women, both implicitly and explicitly.
3. *Gender Prototypical Stimuli:* Implicit stereotypes will be moderated by gender prototypical facial appearance: feminine women and masculine men (i.e., high prototypical targets) should elicit stronger gendered science trait associations than masculine women and feminine men (i.e., low prototypical targets).

4. *Gender Ideology:* Perceiver gender ideology will moderate both implicit and explicit gender-science stereotypes. In particular, gender blindness will relate to weaker stereotypes, and segregationism and assimilationism to stronger stereotypes. There may be a stronger relationship between gender ideology and explicit stereotypes, as both are explicit and more susceptible to controlled processes.
5. *Implicit-Explicit Gender-Science Stereotype Correlation.* There was not a strong hypothesis regarding whether implicit and explicit stereotypes would be related to one another. Previous research has found that the typical correlation between the two ranges from a zero to medium sized correlation (Hofmann et al., 2005).

CHAPTER 4. Study 2 Pre-Test of Scientist Traits

The first step in examining the interface between gender and science stereotypes was to validate the trait content of scientists. Previous research suggests that scientists are viewed as nerdy, odd, peculiar, fun-averse, socially isolated and awkward, unskilled in relationships, obsessed with technology, innately intelligent, unattractive, and singularly focused on their work (Cheryan et al., 2011; Kendall, 1999; Leslie, Cimpian, Meyer, & Freeland, 2015; Losh, 2010; Margolis, Fisher, & Miller, 2000; Rahm & Charbonneau, 1997; Schott & Selwyn, 2000). High school students judged peers who favored science over the humanities as more intelligent and motivated as well as less physically attractive, creative, emotional, socially competent and integrated (Hannover & Kessels, 2004).

In other research, Diekman and colleagues (2010, 2011) showed that compared to medicine, science was implicitly associated with being alone (e.g., *individual, solitary*) rather than together (e.g., *community, group*), and with power (e.g., *achievement, ambition*) rather than warmth (e.g., *affection, nurturing*) (Diekman et al., 2011). STEM careers were also perceived as less likely to fulfill communal goals (e.g., helping others, serving humanity, working with people) relative to both non-STEM male-stereotypic careers (e.g., lawyer, dentist) and female-stereotypic careers (e.g., social worker, nurse). Building upon this research, the present work sought to identify a comprehensive set of both positive and negative traits seen as stereotypic of scientists, determine whether these traits were differentially associated the genders, and examine potential moderators of this gender-science stereotype.

The purpose of the Scientist Trait Pre-test was to validate both positive and negative trait stereotypes of scientists in the scientific domain (i.e., scientific traits and unscientific traits) and warmth domain (i.e. warm traits and cold traits). To do so, gender stereotypes were culled from the warmth and competence literature (Eagly & Steffen, 1984; Fiske et al., 2002), but intentionally selected to represent scientist stereotypes and—as a female-dominated career for comparison—early childhood educator stereotypes (ECE). The goal was to settle on a set of 20 traits in both the science and warmth domain—half of which were positive and half of which were negative—that differentiated perceptions of scientists from ECE. Accounting for valence was important because it allows one to assess whether the groups are viewed differently on semantic content over and above evaluation (Park & Judd, 1990).

Thus, traits selected for scientists were those high on the *scientific* and *cold* dimensions; traits selected for ECE were high on *unscientific* and *warm* dimensions. Within each domain (science vs. warmth) one of these dimensions was positive in valence, and one was negative in valence, permitting a separation of group stereotypes from out-group evaluation. [3] Participants were asked to rate the extent to which the traits described the typical scientist as well as the typical early childhood educator (view all traits in Appendix B).

Participants

Fifty participants (59% female, 80% White, *M* age = 35.16, *SD* = 14.18) were recruited from Amazon.com's Mechanical Turk and paid $.25 each for their participation.[4]

Procedure

Participants were told that we are interested in understanding the characteristics viewed as typical of people working within certain careers, and that they would be randomly assigned to consider 2 of 10 possible careers. Then all were asked to evaluate *scientists* and *early childhood educators (preschool teachers, elementary school teachers)* (ECE) in a counterbalanced order. First, participants wrote a few sentences on what the group is like. They then considered 52 different traits or characteristics presented in randomized order, and

[3] The way traits were selected meant that being high or low on a dimension was confounded with valence, such that being high on the scientific and warmth dimension was positively valenced, and being low on these same dimensions was negatively valenced.

[4] One participant who failed more than 2 attention checks was removed. Attention check questions asked participants to indicate the extent to which traits such as "human", "alive", "breathing", "living", etc. described scientists and teachers; answering below the mid-point of the Likert scale to any of these items was considered failing.

indicated how much each describes or characterizes the target group (1 = Not at all like scientists [early childhood educators], 7 = Very much like scientists [early childhood educators]). After rating both groups, participants judged the favorability of all the traits (1 = Very Unfavorable; 4 = Neither Unfavorable or Favorable; 7 = Very favorable) before completing demographics and being debriefed.

Results

Traits selected to describe scientists were indeed viewed as significantly more typical of scientists than early childhood teachers (except for “capable”). Testing against the mid-point of the scale, each trait index was perceived as significantly favorable or unfavorable as predicted (M scientific = 5.59, SD = .81; M unscientific = 2.10, SD = .99; warm = 6.41, SD = .67; cold = 2.51, SD = 1.04). A 2 (Domain: Science vs. Warmth) × Valence repeated measures ANOVA examining favorability revealed that positively valenced traits were rated more favorably than negatively valenced traits, $F(1, 48) = 404.01$, $p < .001$. Moreover, warm domain traits (warm and cold) were rated more favorably than scientific domain traits (scientific and unscientific) on average, $F(1, 48) = 86.48$, $p < .001$. The lack of a Domain × Valence interaction indicated that the favorability difference between scientific and unscientific traits was not significantly different than that between warm and cold, $F(1, 48) = 2.91$, $p = .09$.

Five of the strongest stereotyped traits (i.e., those that most differentiated scientists from ECE) from each career were selected from within favorable and unfavorable traits to be used in Study 2 (see Table 9). Only four traits were selected for unscientific traits, as these traits did not differentiate between scientists and ECE as well (only gullible showed a significant different between the two, $p < .001$; teachers were seen as marginally more naïve than scientists, $p = .08$; although in the expected direction, ditzy and forgetful showed non-significant differences between ECE and scientists, ps > .26).

For each career, the traits were averaged to create four indices. Alphas were acceptable for scientists (.61, .74, .81, and .68 for scientific, unscientific, warm, and cold, respectively) and teachers (.53, .69, .85 and .76. for scientific, unscientific, warm, and cold, respectively). Using these indices, a 2 (Target: Scientist vs. Teacher) × 2 (Domain: Science vs. Warmth) × 2 (Valence: Positive vs. Negative) × 2 (Participant Gender: Male vs. Female) mixed ANOVA, with repeated measures on the first three factors, was conducted.

	Scientist Traits		**Early Childhood Teacher Traits**	
	Positive (Scientific)	Negative (Cold)	Positive (Warm)	Negative (Unscientific)
	Analytical**	Critical**	Caring**	Gullible*
	Objective**	Isolated**	Compassionate**	Naïve^{+}
	Meticulous**	Robotic**	Gentle**	Forgetful
	Logical*	Self-absorbed**	Nurturing**	Ditzy
	Scientific**	Unsociable**	Warm**	

Table 9. Traits that most differentiated scientists from ECE in the pre-test. ***p* < .001. **p* < .01, $^{+}$*p* < .10.

Supporting Hypothesis 2, a large three way interaction (Target × Domain × Valence) confirmed there were strong trait stereotypes differentiating scientists from ECE, $F(1, 47) = 444.35$, $p < .001$. This three-way interaction captured stereotypes that scientists are more associated with scientific and cold traits than teachers, who are more associated with unscientific and warm traits than scientists. Another way of thinking about the interaction is that scientists are seen as more scientific than unscientific, and more cold than warm, relative to teachers. As seen in Figure 5, each stereotype index (i.e., the simple effect of target – scientists vs. ECE – within each trait index) confirmed that scientists were seen as significantly more scientific ($F(1, 47) = 238.82$, $p < .001$) and cold ($F(1, 47) = 185.01$, $p < .001$) than ECE; whereas ECE were seen as significantly more warm ($F(1, 47) = 279.46$, $p < .001$) and unscientific ($F(1, 47) = 9.15$, $p < .01$) than scientists (see Figure 5). There were no effects involving participant gender on any of the stereotype indices, all Fs(1, 47) ≤ 0.42, ps ≥ .52.

Scientist Trait Pre-Test Discussion

The pre-test successfully identified traits that differentiated scientist and ECE, although this differentiation was markedly smaller for unscientific traits relative to the other stereotype indices. This may be due to a general reluctance to ascribe any negative traits to both ECE and scientists, but especially to ECE (a sizable Target × Valence interaction confirmed that people ascribed more positive traits overall to ECE than to scientists regardless of trait domain, $F(1, 47) = 48.03$, $p < .001$). In fact, of all the traits tested in the pre-test, not a single negative trait was rated as significantly descriptive of ECE, and only 2 were seen as significantly descriptive of scientists, *isolated* and *critical*.

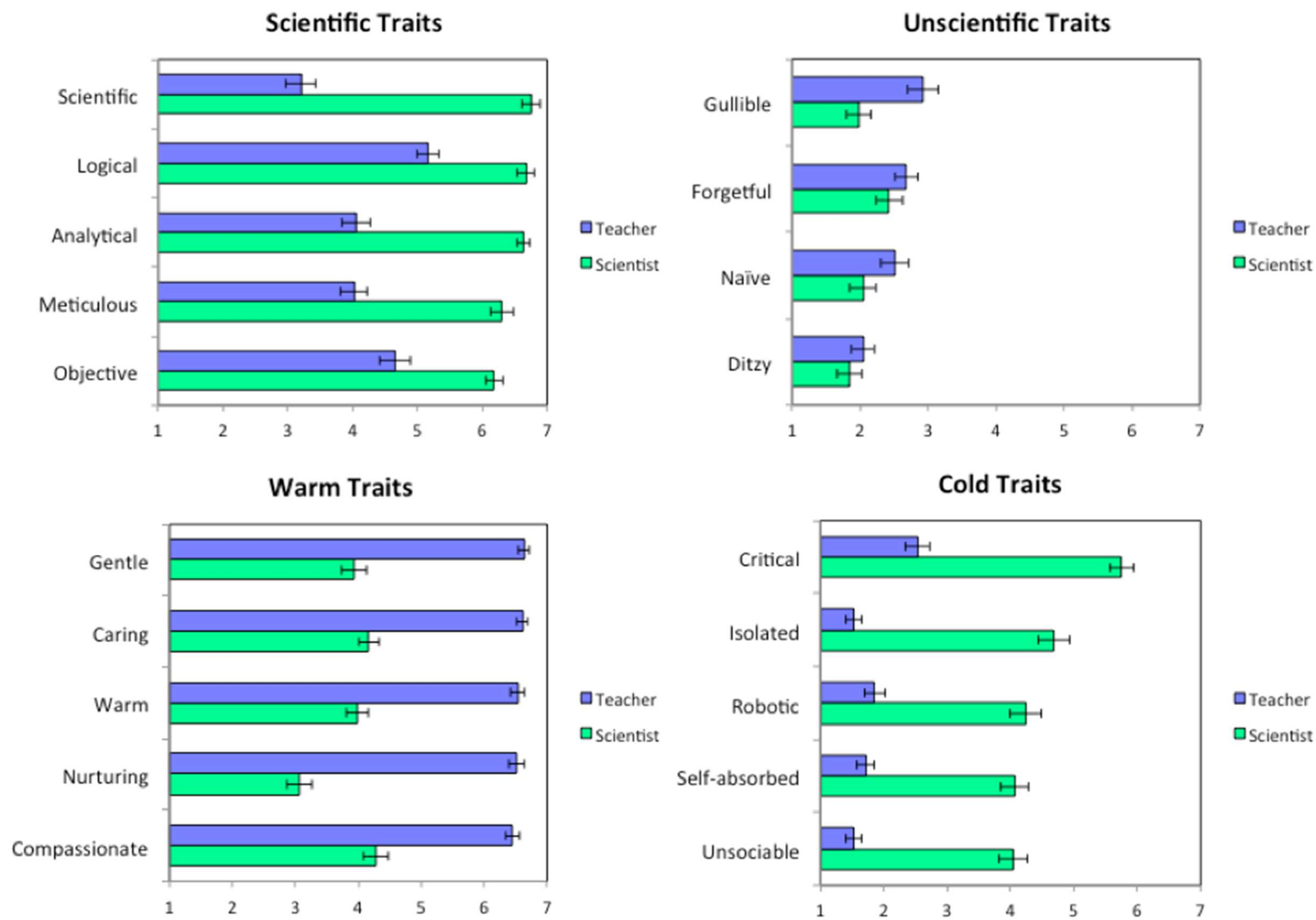


Figure 5. Average endorsement of each trait for Scientist and ECE.

CHAPTER 5. Study 2 Proper

Having validated a set of traits that captured scientist stereotypes, the goal of Study 2 Proper was three-fold: to examine 1) how these traits were differentially ascribed to men and women at both an implicit and explicit level (i.e., to measure *gender-science stereotypes*), 2) whether gender ideology moderates implicit and explicit gender-science stereotypes, and 3) whether gender prototypicality moderates implicit gender-science stereotypes. I hypothesized that both implicit and explicit stereotypes would emerge (Hypothesis 2), that stereotypes would be moderated by target gender prototypicality (Hypothesis 3, tested only for implicit stereotypes), and that the both implicit and explicit stereotypes would be moderated by perceiver gender ideology (Hypothesis 4).

Both implicit and explicit gender-science stereotypes were examined because literature shows that they may reflect different processes and have different consequences for behavior. Although implicit attitudes have been defined multiple ways (Bargh, 1989, 1994)[5], there is general agreement that they operate via automatic processes that are efficient, occur without awareness, are difficult to control, and are therefore unintentional (for reviews, see Banaji, 2001; Blair, 2001; Devine & Monteith, 1999; Fiske, 1998, but see Hahn, Judd, Hirsh, & Blair, 2014, for evidence that people are surprisingly aware of their implicit attitudes). In contrast, explicit attitudes operate via non-automatic processes that are less efficient (i.e., slower), occur with awareness, are controllable, and are therefore intentional (Bargh & Chartrand, 1999; Nosek, 2007). Unlike implicit attitudes, explicit attitudes are modified based on conscious thoughts or propositions; although people may possess implicit associations of men as more mathematical than women, conscious propositions (e.g., “I know many women who are good at math, and quite a few men who aren’t good at math” or “I don’t want to be sexist”) can result in a more egalitarian explicit attitude (Gawronski & Bodenhausen, 2007). Moreover, participants may not

[5] I define attitudes as encompassing both stereotypes and evaluations.

be willing or comfortable expressing explicit beliefs that are socially undesirable, and/or they may not have full access to their attitudes. Indeed, despite the pervasive tendency to implicitly associate men more than women with math (Nosek et al., 2009), most people do not explicitly endorse that men are better at math than women (Hyde, Fennema, Ryan, Frost, & Hopp, 1990; Schmader et al., 2004).

Supporting the distinctiveness of implicit and explicit attitudes, research shows that the relationship between them for a given attitude object varies greatly in size, ranging from a zero to a medium correlation depending on a number of additional moderators, such as the degree of conceptual correspondence between the two measures (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, 2005, 2007). For example, Nosek and colleagues found only a modest correlation (r = .42) between women's implicit gender-math stereotypes (i.e., math = male) and their explicit math = male stereotypes (Nosek, Banaji, & Greenwald, 2002), again suggesting that implicit and explicit stereotypes are somewhat distinct constructs. And whereas explicit attitudes are more strongly related to clearly controlled behaviors (e.g., what is said during a conversation), implicit attitudes are more strongly related to subtle, automatic and non-conscious behaviors (e.g., non-verbal behavior; Dovidio, Kawakami, & Gaertner, 2002).

For these reasons, it was worthwhile to measure both implicit and explicit gender-science stereotypes. A go/no-go association task (GNAT; Nosek & Banaji, 2001) was used to assess implicit gender-science stereotypes.[6] The GNAT consists of quickly presenting participants with items from a number of categories and asking them to press the spacebar ("go") whenever an item from one of two focal, target categories (e.g., "men" and "scientific") appeared and to make no response ("no-go") when the item was not from these two categories (e.g., "women", "unscientific", and typically some irrelevant background category, in this case

[6] The most commonly used implicit measure, the Implicit Association Task (IAT), was avoided because it asks participants to categorize two social groups at the same time, creating a forced comparison between two target groups. The GNAT, on the other hand, allows one to assess associations towards one group at a time, with the other social group (and other distractor stimuli) in the background.

images and names of birds). The response deadlines used in the task are very fast, typically between 500-1000 ms, making strategic, controlled responding very difficult. The easier the task is for participants (i.e., the better they perform), the stronger the association or fit between the two focal categories. Variation in responses is largely due to the error rates, and d’ (Green & Swets, 1966) is used as a measure of sensitivity to the difference between focal and non-focal categories. A higher d’ indicates an easier time simultaneously keeping the two focal categories in mind and distinguishing them from background categories.

In addition to establishing the existence of gender-science stereotypes both at an implicit and explicit level, I examined whether they depended on participant gender ideologies. Specifically, I expected gender blindness to be related to weaker stereotypes, and assimilationism, segregationism, and potentially gender awareness, to be related to stronger stereotypes. Research in the ethnic and racial domain has shown that ethnic ideologies can act as important moderators of both implicit and explicit racial stereotyping and prejudice. For example, Richeson and Nussbuam (2004) demonstrated that inducing a colorblind ideology resulted in greater implicit (and explicit) pro-White racial bias relative to inducing a multicultural ideology. A weaker implicit pro-White bias has also been found among those who possess more positive explicit racial attitudes towards Blacks and those who are high on internal (and low on external) motivation to respond without prejudice (Devine, Ashby, Amodio, Harmon-Jones, & Vance, 2002; McConnell & Leibold, 2001). A final goal was to examine how implicit stereotype strength might depend on the target’s gender prototypical facial appearance (i.e., the target’s feminine vs. masculine facial appearance), and how this might interact with the perceiver’s gender ideology.

Design

The general outline of the study is as follows. First, key explicit measures of interest—gender ideology, gender-science stereotypes, gender evaluations, and gender-STEM stereotypes—were collected at the beginning of the semester during a large online pre-screen (called Study 2 pre-screen). Participants who completed the pre-screen were invited to participate in the laboratory study later in the semester, during which their implicit gendered trait stereotypes and evaluations were assessed (Study 2 lab session). Collecting explicit measures in the pre-screen enabled a separation of explicit judgments from implicit judgments so that they did not interfere with one another.

In the lab session, participants completed a GNAT, which assessed their implicit associations of men and women as scientific, unscientific, warm, cold, good, and bad. Good and bad associations were included to examine whether the moderators of interest differentially affected stereotyping vs. evaluation. Critically, participants were randomly assigned to view one of three different kinds of gender stimuli as they completed the implicit task: generic gender names, high prototypical images of men and women (i.e., feminine women and masculine men), or low prototypical images of men and women (i.e., masculine women and feminine men; pre-tested in Study 1). The study was thus a 2 (Target Gender: Men vs. Women) × 2 (Trait Domain: Science vs. Warmth) × 2 (Trait Valence: Positive vs. Negative) × 2 (Participant Gender: Male vs. Female) × 3 (Stimuli Condition: Generic, Low Prototypic, High Prototypic) mixed factorial design with the first three factors manipulated within subjects and the last factor manipulated between subjects. After completing the GNAT, participants completed the same explicit measures included in the pre-screen, as well as a couple additional measure of interest that could not be included in the pre-screen due to length restrictions. Explicit judgments were collected both during lab session in case an adequate number of participants could not be recruited, or if a substantial number of participants could not be matched to their pre-screen data.

Participants

Participants were 372 students at the University of Colorado Boulder (212 women, 160 men; 73% White, 11.3% Asian, 4.3% Latino, .8% Black, and 10.6% biracial; mean age = 19.29, *SD* = 2.51). All participants were recruited through the Psychology Department subject pool and had voluntarily completed the mass pre-screen at the beginning of the semester; 360 were matched to their pre-screen data using their self-reported participant ID code. Participants received partial course credit in exchange for their participation.

Materials

Pre-screen

See Appendix C for the complete set of all explicit measures included in the pre-screen.

Gender Ideology.

The 18-item Gender Ideology scale assessed the four gender ideologies (see Table 8 for all items), on a 1 to 7 scale (Strongly Disagree to Strongly Agree). In the first semester collection of the pre-screen, all items were presented in a randomized order on the same page; in the second semester, items were presented on separate pages with 5-6 items randomly pre-selected for each page.[7]

Explicit Gender-science Stereotypes.

The percent estimate task (Park & Rothbart, 1982) asked participants to think about women [men] in the United States and estimate the percentage of each group that possessed the pre-tested traits stereotypic of scientists. Only four traits from each index were included because of pre-screen length limitations (see Table 10). Because the pre-test on traits was largely unsuccessful in selecting unscientific traits that differentiated scientists and ECE, “Forgetful” was replaced with “Incompetent,” and “Gullible” was replaced with “Imprecise”, traits that we believed had higher face validity in assessing the concept *unscientific*. In retrospect, “Gullible” should have been retained in the trait estimation task, as it differentiated scientists and

[7] In the second semester pre-screen, several exploratory gender ideology items were included. Using separate pages and this randomization strategy ensured that they did not appear on the same page as the original items.

ECE in the Scientific Trait pre-test.[8] Due to design limitations in the pre-screen survey, women were always evaluated first, followed by men. For each trait, participants were asked to type in a number between 0 and 100; responses outside these bounds were treated as missing (*N* = 3).

Scientist Traits		**ECE Traits**	
Positive (Scientific)	Negative (Cold)	Positive (Warm)	Negative (Unscientific)
Analytical	Critical	Caring	Ditzy-
Objective	Isolated	Gentle	Incompetent*
Logical	Robotic	Nurturing	Imprecise*
Scientific	Self-absorbed	Warm	Naïve

Table 10. Traits presented in the pre-screen. *Trait not included in pre-test.

Explicit Evaluation.

A feeling thermometer assessed warmth towards men and women on a 100-point scale (0 = “cool and unfavorable feelings”, 100 = “warm and favorable feelings”). A graphic of the thermometer scale was provided (Figure 6). Target groups were presented in randomized order on the same page. Participants were asked to type in their responses from 0 to 100, and numbers beyond these bounds (*N* = 2) were treated as missing.

Figure 6. Feeling thermometer visual presented to participants in the pre-screen.

[8] The opposite of “incompetent”, “competent”, was included in the pre-test, and scientists were rated as significantly more competent than ECE. Neither *precise* nor *imprecise* were pre-tested.

Gender-Science Stereotype.

Two gender-STEM stereotype items assessed self-stereotypes *("According to my own personal beliefs, I generally expect men to do better in math and science than women"*) and societal stereotypes *("According to general beliefs in society, men are expected to be better at math and science than women"*), from 1 (Strongly Disagree) to 7 (Strongly Agree).

Lab Session

GNAT.

The Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001) assessed implicit science trait stereotypes and implicit gender evaluations. Each participant completed a set of twelve GNAT blocks that assessed their implicit stereotypic associations of men and women in both the science (scientific and unscientific traits) and warmth domains (warm and cold traits), as well as their implicit evaluative associations of men and women (good and bad words). Eight focal categories of stimuli were developed to assess these: men and women; traits representing scientific, unscientific, warm and cold; and gender-neutral nouns representing good and bad. One additional category, birds, was also included to increase variability in the distractor category. Scientific, unscientific, warm and cold categories were each instantiated by the five traits that most strongly differentiated between scientists and ECE in the pre-test.[9] "Good Words" and "Bad Words" were nouns that were selected to be gender-neutral (see Table 11). Recall that within a given block of the GNAT, participants were quickly (i.e., 600ms) presented with exemplars of men and women, intermixed with words (as well as birds).

[9] Two traits included in the Study 2 pre-screen to assess unscientific traits were not included in the GNAT—*incompetent* and *imprecise*. By the time of the proposal defense, the pre-screen was already underway, and committee members were concerned that these traits would be difficult to process during the GNAT task (because they are negated versions of other words. A novel trait, "careless", was added to the unscientific dimension in the GNAT.

For a given block, they were instructed to “go” (press the space-bar) to any stimuli that belonged to one of two categories—either men or women and traits or words from one of the stereotype (e.g., scientific) or evaluative (e.g., bad) categories. Thus participants completed 12 GNAT Blocks in total: each of the six categories of trait or evaluative word stimuli was paired once with women and once with men.

Category Labels						
GOOD WORDS	**BAD WORDS**	**SCIENTIFIC TRAITS**	**UNSCIENTIFIC TRAITS**	**WARM TRAITS**	**COLD TRAITS**	**BIRDS**
BEACH	COCKROACH	ANALYTICAL	DITZY	CARING	CRITICAL	BLUEBIRD
PARADISE	DISEASE	LOGICAL	CARELESS	COMPASSIONATE	ISOLATED	BUNTING
PUPPY	FILTH	OBJECTIVE	FORGETFUL	GENTLE	ROBOTIC	CARDINAL
SMILE	POISON	METICULOUS	GULLIBLE	NURTURING	SELF-ABSORBED	EAGLE
SUNRISE	VOMIT	SCIENTIFIC	NAÏVE	WARM	UNSOCIABLE	LARK
						OWL
						PELICAN
						ALBATROSS
						STARLING
						SWALLOW

Table 11. Category labels and stimuli presented in the GNAT.

To examine the impact of gender prototypicality on stereotypes and evaluations, different instantiations of the categories “men” and “women” were employed: generic stimuli (common gender names, *n* = 126), high prototypic gender photos (i.e., feminine women and masculine men, *n* = 124) or low prototypic gender photos (i.e., masculine women and feminine men, *n* = 122). In the Generic Names Condition, the stimuli representing men and women were common gender names equated in their length (i.e., Jason, David, Brian, Kevin, Andrew, Robert, Stephen, Jeffrey, Michael and Benjamin; Sarah, Laura, Susan, Julia, Nicole, Amanda, Heather, Rebecca, Shannon and Michelle; Park et al., 2010). In the High Prototypic Gender Stimuli Condition, the stimuli consisted of the ten most feminine women and ten most masculine men from Study 1; in the Low Prototypic Gender Stimuli Condition, the stimuli consisted of the ten most masculine women and ten most feminine men from Study 1. All faces were cropped to only be from the shoulders up. Each was then placed on the same sized white background to standardize the size in which they would be displayed (see Figures 7 and 8).

Figure 7. All women presented in the GNAT. The top 2 rows were used in High Prototypic Condition, and the bottom 2 were used in the Low Prototypic Condition.

Figure 8. All men presented in the GNAT. The top 2 rows were used in High Prototypic Condition, the bottom 2 were used in the Low Prototypic condition.

The design of the GNAT was thus a 2 (Target Gender: Men or Women, within subjects) × 3 (Domain: Science Traits, Warmth, or Evaluative, within subjects) × 2 (Valence: Positive vs. Negative Traits/Words, within subjects) × 3 (Gender Stimuli Type: Generic names, High Prototypic Photos, Low Prototypic Photos; between subjects) mixed design, with repeated measures on the first three factors. Each participant completed all 12 GNAT blocks in one of 12 counterbalanced orders (see Appendix E). Across the twelve blocks, they were instructed to “go to” the following stimuli and ignore all other stimuli:

1. Men and “Scientific Traits” (i.e., positive scientist triats)
2. Women and “Scientific Traits”
3. Men and “Cold Traits” (i.e. negative scientist traits)
4. Women and “Cold Traits”
5. Men and “Warm Traits” (i.e., positive early childhood educator traits)

6. Women and “Warm Traits”
7. Men and “Unscientific Traits” (i.e., negative early childhood educator traits)
8. Women and “Unscientific Traits”
9. Men and “good words” (i.e., gender-neutral positive nouns)
10. Women and “good words”
11. Men and “bad words” (i.e., gender-neutral negative nouns)
12. Women and “bad words”

Each block contained 100 trials (i.e., presentation of different stimuli), 80 of which represented the focal categories of interest, and 20 of which were distractor trials with bird names or images (see Table 12 for details on the stimuli presented in each GNAT task and block).

	Science Domain Stereotypes				Cold Domain Stereotypes				Evaluation			
	F + S	M + S	F + US	M + US	F + Warm	M + Warm	F + Cold	M + Cold	F + Good	M + Good	F + Bad	M + Bad
Female Stimuli	**20**	20	**20**	20	**20**	20	20	20	20	20	20	20
Male Stimuli	20	**20**	20	**20**	20	**20**	20	20	20	20	20	20
Scientific traits	**20**	**20**	20	20								
Unscientific traits	20	20	**20**	**20**								
Warm traits					**20**	**20**	20	20				
Cold traits					20	20	**20**	**20**				
Good words									**20**	**20**	20	20
Bad words									20	20	**20**	**20**
Bird Stimuli	20	20	20	20	20	20	20	20	20	20	20	20

Table 12. Frequency of stimuli presentation per GNAT block. S=Scientific, US = Unscientific. In each block (column) 20 trials from each of the 2 focal categories (e.g., 20 trials with female names, 20 trials with scientific words), 20 trials from each of the 2 non-focal categories (e.g., 20 trials with male names, 20 trials with uncientific words), and 20 trials from the distractor category (birds) make a total of 100 trials per block.

For example, for Block 1 above, 20 trials contained *male* stimuli (to which participants should “go”), 20 contained *female* stimuli (“no-go”), 20 contained *scientific traits* (“go”), 20 contained *unscientific traits* (“no-go”), and 20 contain bird stimuli (“no-go”). For participants in the Generic Names Condition, bird stimuli were also names; for participants in the Low Prototypic or High Prototypic Gender Stimuli Conditions, bird stimuli were also photos. Each of the gender stimuli was presented twice, and each of the word stimuli was presented four times. Stimuli were presented for a 600 ms response window with an inter-stimulus interval of 150 ms. Error feedback (a large red “X”) was provided, and a blue “!!” appeared if participants were too slow to respond. A correct response simply advanced to the next stimulus. The labels for the task were presented in light blue against black, and the stimuli were presented in black on a white rectangle in the center of the screen, where they remained until the participant responded or a response deadline was reached.

Four practice blocks were also developed for participants to become familiar with the GNAT before beginning experimental blocks. Each included 40 stimulus presentations, half of which were to target stimuli. The first practice block consisted of responding to a single category at a time, starting with *birds* (distractor stimuli of men and women; response deadline of 1000ms), then *women* (distractor of men and birds, response deadline of 800ms), and *men* (background of women and birds, response deadline of 800ms). The fourth practice block had a shorter response deadline, and involved responding to birds or cats (distractors were men and women, response deadline of 600ms).

Gender stimuli presented in the practice trials varied depending on Gender Stimuli Condition—participants in the Generic Names Condition completed practice trials with generic names, whereas those in the Low Prototypic or High Prototypic Gender Stimuli Conditions saw the same photos they would view in the study proper. Gender stimuli were also included as distractors to increase familiarity with the stimuli that would be encountered during experimental blocks.

Explicit Measures.

Explicit measures in the laboratory session largely overlapped with those included in the pre-screen. Notably, in the laboratory explicit measures, all traits from both the pre-test and the pre-screen were included (i.e., for unscientific, all 6 traits, plus an additional trait, “careless”, were used), and thermometer ratings included not only the targets women and men, but also specific subgroups of men and women representing traditional vs. nontraditional gender roles: female and male politicians, female and male scientists, female and male nurses, and stay at home moms and dads.[10] Additional explicit measures included in the lab session, the ambivalent sexism inventory (Glick & Fiske, 1996) and the internal vs. external motivation to respond without sexism scale (Klonis, Plant, & Devine, 2005), are described in the ancillary analysis section.

Procedures

Pre-screen

Participants volunteered to partake in a large online battery of survey measures conducted at the beginning of each semester. During the online pre-screen, participants completed measures in the following order: 1) gender ideology scale, 2) trait estimation task for *women*, followed by *men*, 3) thermometer ratings of men and women, and finally 4) personal and perceived societal endorsement of gender-STEM stereotypes. Measures in the pre-screen were intermixed with other measures included by other researchers. All measures were presented on separate pages, and all items within each measure were randomized for each

[10] Three to four target groups were pre-selected to be presented on three separate pages to ensure that the same two targets of different genders were not presented on the same page. Gender ideology items were all presented on a single page in randomized order. Unlike the pre-screen, percent estimate ratings and thermometer ratings were made on a sliding scale from 0 and 100 rather than entering numbers manually.

participant with the exception of general gender-STEM stereotypes—personal stereotypes were always rated before perceived societal stereotypes.

Laboratory Session

Those who participated in the pre-screen were invited to come into the lab in groups of up to 10 people. They were seated at an individual station with a Macintosh laptop computer on which the entire experiment was conducted. After participants read through the informed consent, the experimenter explained the implicit task that they would be completing (see Appendix D for the script). Verbal instructions were reinforced with written instructions provided on each computer. Before beginning the study proper, participants completed four practice blocks of the GNAT at their own pace. Once everyone had finished the practice rounds, the experimenter passed out candy to participants while acknowledging that the task would be challenging, and explained they were ready to begin the experimental trials of the GNAT at their own pace.

Participants were then given time to study all of the gender stimuli (names or the photos, depending on Gender Stimuli Condition), category labels and word stimuli that they would be asked to respond to during the experimental trials. After viewing a list of each category (e.g. “Scientific Words”) and the stimuli within that category (see Figure 9), each stimulus within that category was individually presented one at a time, and participants were instructed to, for each word, think about a person, image, or memory the word brings to mind. These were presented in the following fixed order: Good words, bad words, scientific traits, unscientific traits, warm traits, and cold traits.

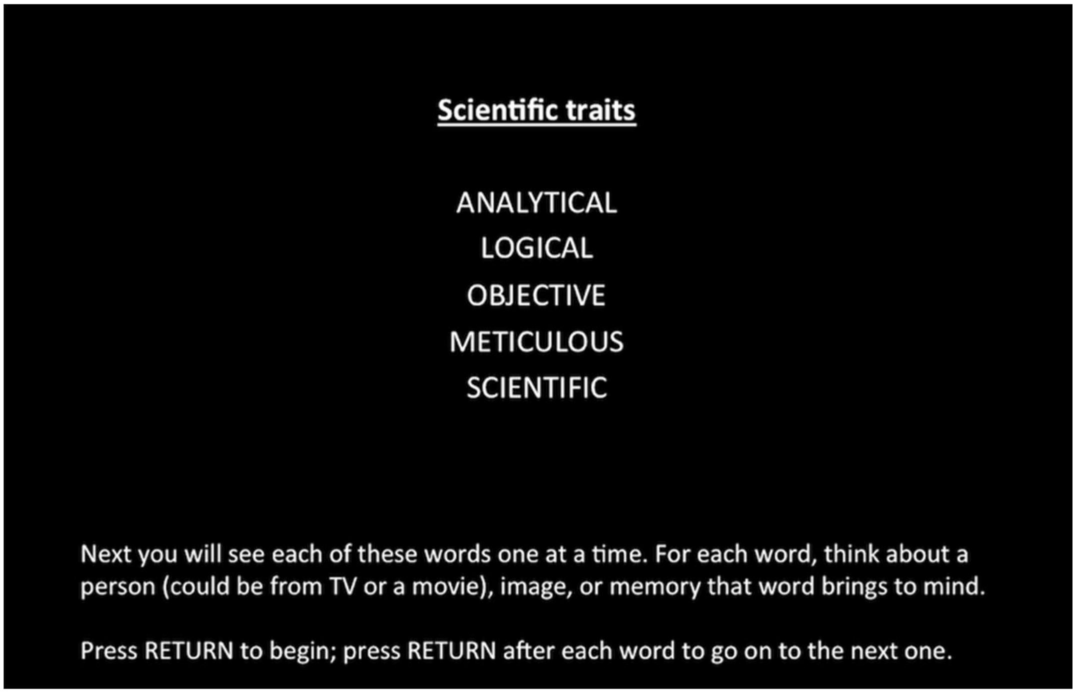


Figure 9. Example presentation of words within each category in the GNAT.

Finally, participants were reminded that they could refer to a piece of paper on their desk that had explicitly labeled each of the categories and the words that belonged within it (see Appendix D). Participants were encouraged to take breaks in between the blocks of the GNAT to stretch or have some candy, but informed that once they began a block they could not take a break. Before each block began, participants were clearly informed of which categories they would be pairing together:

> "In this block of trials, your task is to select things from the following categories:
> WOMEN or WARM traits
> When you see either WOMEN (a photo of a woman) or a WARM trait, you must press the spacebar as quickly as possible while the word or image is on the screen. You should ignore all other things. As you have seen, there is very little time to make each decision. To be accurate, it is important that you concentrate. Ask the experimenter if you have any questions.
> Press RETURN to continue."

Participants had as long as they desired to pause and study the words before starting. Once ready, they pressed the spacebar to begin the block. The "Get Ready!" appeared for 600ms before the block began.

Explicit Measures.

The GNAT took approximately 30 minutes for participants to complete, at which point participants began the survey portion of the study in Qualtrics Survey Software at their leisure. After collecting basic information to match participants to their pre-screen data, measures were presented on separate pages and in the following order: 1) explicit gender-science stereotypes for women and men (target order counter-balanced), 2) thermometer ratings, 3) gender-STEM stereotypes, 4) gender ideology, 5) ambivalent sexism inventory 6) and internal/external motivation to respond without sexism 7). On any given page, items were presented in a randomized fashion.

Finally, participants reported their gender, age, ethnicity, education, and political orientation. Once participants were finished, the read a debriefing form that asked them to remain seated until dismissed. Once everyone was finished the task, all participants were thanked and able to leave.

Results

Wherever possible, explicit measures collected during pre-screen were analyzed rather than measures collected during the laboratory session. Pre-screen measures were more compelling because they were separated in time from the implicit task, which helps to minimize any carryover or interference effects between the two. Some missing responses on explicit measures contribute to varying degrees of freedom.

Results are presented in the following order:

1. Explicit gender-science stereotypes
2. Implicit gender-science stereotypes
    a. Their moderation by gender prototypicality.
3. The impact of perceiver gender ideology on stereotypes
    a. For explicit gender-science stereotypes
    b. For implicit gender-science stereotypes
4. The relationship between implicit and explicit gender science-trait stereotypes
5. Relationships between other explicit measures
6. Ancillary analyses

a. Measurement properties of gender ideology and gender-science stereotypes
b. Implicit and explicit gender evaluation patterns and moderation by perceiver gender ideology

Data Analysis

Like the pre-test that assessed stereotypes of scientists, overall gender-science stereotypes were captured by the same three-way interaction: 2 (Target: Men vs. Women) × 2 (Trait Domain: Scientific and Unscientific vs. Warm and Cold) × 2 (Trait Valence: Positive vs. Negative). This three-way interaction captures associating men more than women with *scientific* traits and *cold* traits, and women more than men with *unscientific* and *warm* traits. Another way of interpreting this effect is that the difference between scientific and unscientific is greater for male targets than female targets (that a large percentage of males are seen as having scientific traits and a small percentage as having unscientific traits, whereas this differences is smaller for females), and simultaneously the difference between warm and cold being greater for female targets than male targets (that a large percentage of females are seen as having warm traits and a small percentage as having cold traits, whereas this differences is smaller for males). I call this index *overall stereotype strength.*

I further broke this index down by trait dimension in order to examine where the genders were perceived differently. Thus I computed a stereotype index for each trait dimension, coded in the stereotypic direction so that higher numbers reflect more stereotypic perceptions (i.e., men scientific – women scientific; men cold – women cold; women warm – men warm; women unscientific – men unscientific). In all analyses, participant gender was included as a factor unless otherwise noted.

Explicit Gender-science Stereotypes

I begin with a test of *Hypothesis 2*: *people possess explicit stereotypes differentially attributing science stereotypic traits to men and women* (i.e., gender-science stereotypes). The percent of men and women estimated to have each trait was examined as a function of target gender, domain, and valence in order to assess explicit gender-science stereotypes (see Table 13 for means and alphas, as well as Figure 10).

Index	*M* (*SD*)	Alpha	*N* responses	*N* Items
Women Scientific	50.91 (14.76)	.68	340	4
Men Scientific	56.89 (13.84)	.74	338	4
Women Unscientific	32.31 (15.20)	.77	340	4
Men Unscientific	32.23 (13.76)	.72	337	4
Women Warm	69.64 (13.87)	.83	342	4
Men Warm	45.41 (17.15)	.89	337	4
Women Cold	39.41 (13.34)	.60	341	4
Men Cold	44.31 (13.51)	.62	338	4

Table 13. Descriptive statistics for explicit gender-science stereotypes.

The mean trait ratings were examined in a 2 (Target Gender: Men vs. Women) × 2 (Domain: Science vs. Warmth) × 2 (Valence: Positive vs. Negative) × 2 (Participant Gender: Male vs. Female) mixed ANOVA with repeated measures on the first three factors. Of primary interest was the Target Gender × Domain × Valence interaction indicating that there were strong gender-science stereotypes, $F(1, 333) = 476.11$, $p < .001$. Looking at each index individually, the target effect was significant for scientific traits, $F(1, 335) = 66.14$, $p < .001$, indicating that men were judged as more scientific than women. Unexpectedly, there was no target difference for unscientific traits, $F(1, 334) = .01$, $p = .90$. Participants possessed very strong warmth stereotypes, with women judged as warmer than men, $F(1, 334) = 733.06$, $p < .001$. Men were also judged as colder than women, $F(1, 355) = 75.53$, $p < .001$ (see Figure 10).

The-three way interaction capturing stereotypic perceptions of the genders on these scientist related traits was moderated by participant gender, $F(1, 333) = 16.06$, $p < .001$, which indicated that the effect was weaker among females than among males, although strong within both genders, *Fs*(1, 333) ≥ 187.14, *p*s < .001 (see Figure 11).

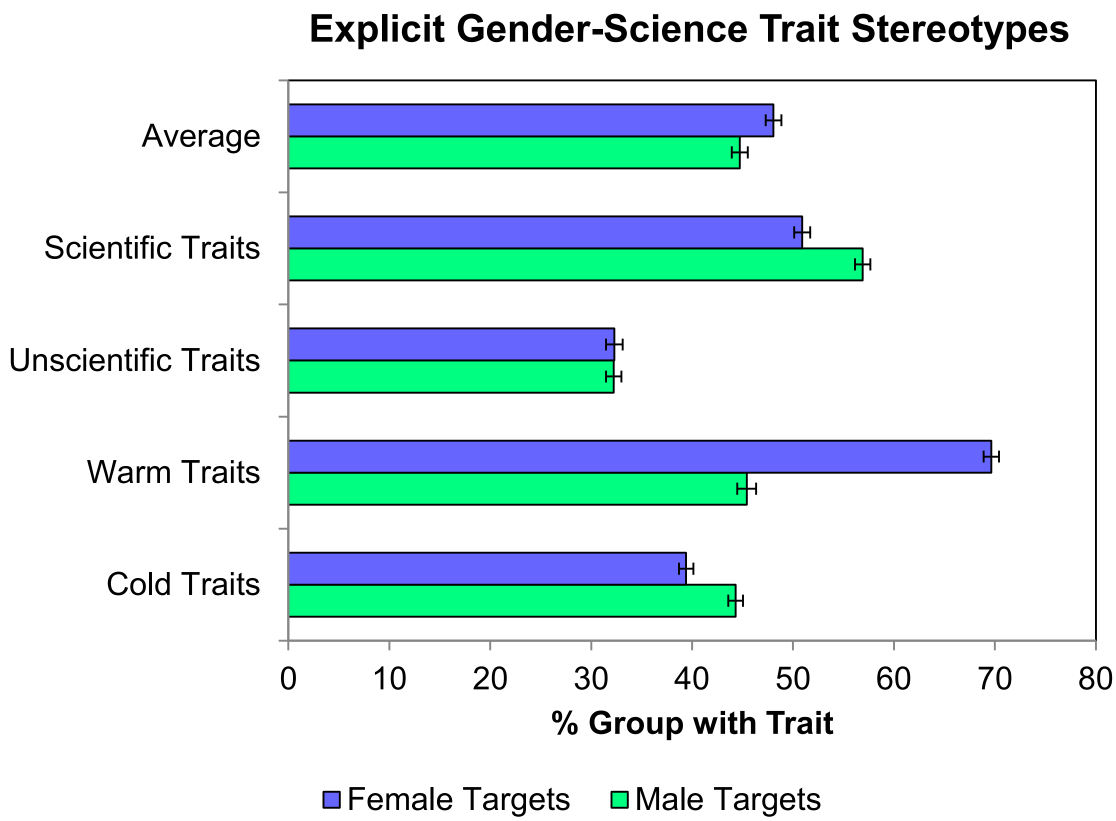


Figure 10. Explicit gender-science stereotypes by target gender.

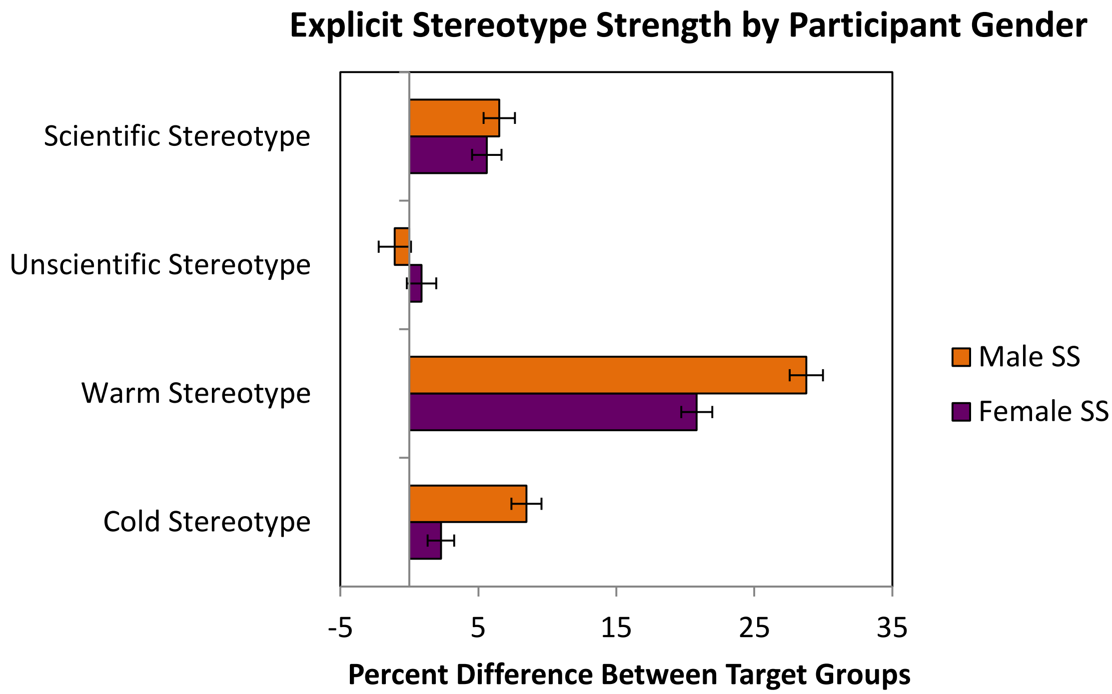


Figure 11. Explicit stereotype strength by participant gender. Male SS = male participants. Female SS = female participants. Stereotype strength was defined as the mean rating for the stereotypic minus counterstereotypic group.

Looking within each trait dimension, there were no participant gender differences for scientific and unscientific traits *F*s(1, 334) ≤ 2.34, *p*s ≥ .13. For both warm and cold traits, however, females expressed weaker stereotypes than males, *F*s(1, 333) ≥ 20.07, *p*s < .001, although they still possessed stereotypes, *F*s(1, 333) ≥ 8.28, *p* < .01. Complete ANOVA results can be viewed in Table 14. There were other interactions with participant gender, but these were all qualified by the 4-way interaction and so are not discussed.

Effect	F	df	p
TGender	81.71*	1, 333	<.001
Domain	246.09*	1, 333	<.001
Valence	570.08*	1, 333	<.001
TGender × Domain	385.51*	1, 333	<.001
TGender × Valence	257.41*	1, 333	<.001
Domain × Valence	45.55*	1, 333	<.001
TGender × Domain × Valence	476.11*	1, 333	<.001
PGender Effects			
PGender	6.84*	1, 333	.01
Domain × PGender	5.27*	1, 333	.02
Valence × PGender	.80	1, 333	.37
TGender × PGender	.07	1, 333	.78
TGender × Domain × PGender	3.91	1, 333	.05
TGender × Valence × PGender	25.17*	1, 333	<.001
Domain × Valence × PGender	.05	1, 333	.82
TGender × Domain × Valence × PGender	16.06*	1, 333	<.001

Table 14. ANOVA output for explicit overall stereotypes. TGender = Target Gender. PGender = Participant Gender. *Significant at $p < .05$.

In summary, the genders were judged in a stereotypic manner on 3 of the 4 dimensions relevant to science stereotypes: men were seen as more scientific and cold, and as less warm than women. The genders were not perceived differently on unscientific traits, however. It is also striking that, as seen in Figure 10, stereotypes were much stronger in the warmth domain (i.e., for warm and cold traits) than in the science domain (i.e. for scientific and unscientific traits). Although both males and females endorsed explicit gender-science stereotypes, female participants had weaker stereotypes for both warm and cold traits.

Men and women were not seen differently on *unscientific* traits. This may reflect a reality in which men and women are viewed as equal on this dimension; however, it is also the case that the items used in the pre-test on which scientists were seen as significantly higher than ECE were slightly different than those used in the pre-screen, where stereotypes of the genders were assessed (*forgetful* and *gullible* were used in the pre-test; these were replaced with *incompetent* and *imprecise* in the pre-screen). In the ancillary analysis section, I show that the unscientific stereotype does in fact emerge in explicit judgments made during the laboratory session (whether defined by the traits used in the pre-test or those used in the pre-screen), suggesting that although gendered science stereotypes are weaker for this dimension, they still appear to exist.

Implicit Gender-science Stereotypes

Hypothesis 2 also maintained that in addition to explicit gender-science stereotypes, *people possess implicit gender-science stereotypes*. To assess the strength of implicit stereotypes, responses to each of the 12 GNAT blocks were analyzed by calculating d' (the proportion of Hits to the focal categories (correct *Go* response [minus the proportion of False Alarms to the background categories [incorrect *Go* response], after first transforming these proportions to their respective *z* values from the standard normal distribution).[11] Extreme cell values (0 or 1) were corrected following recommendations by Banaji and Greenwald (Banaji & Greenwald, 1995). If a person had 0 hits or false alarms in a given block, their score was changed to 1/(2*the potential number of hits[false alarms]), or 1/(2*40) for hits and 1/(2*60) for false alarms in the experimental blocks. Alternatively, if a person had a proportion of 1 for either hits or false alarms, their score was transformed to 1 – (1/(2*the potential number of hits[false alarms), or 1-(1/(2*40) for hits and 1-(1/(2*60)) for false alarms. Two participants had excessive errors rates on one or two blocks (i.e., d' < 0); results were the same with them removed, so they were retained in the analysis.

To examine gender stereotypes, a 2 (Target Gender: Men vs. Women) × 2 (Domain: Science vs. Warmth) × 2 (Valence: Positive or Negative) × 2 (Participant Gender: Male vs. Female) × 3 (Gender Stimuli Condition: Generic Names, High Prototypic Gender Stimuli, Low Prototypic Gender Stimuli) mixed ANOVA was conducted, with repeated measures on the first three factors. Between subjects factors were orthogonally contrast coded. The analysis presented here includes Gender Stimuli Condition as a factor, but to facilitate the presentation

[11] A version of the analyses removed birds from the background categories prior to calculating d'. Results were the same, and so birds were included as a background category.

of the results, I will wait to discuss effects involving Gender Stimuli Condition until the next section.

Of primary interest was the Target Gender × Domain × Valence interaction, which captured overall implicit gender-science stereotypes (in the same manner as the explicit trait ratings). This interaction was highly significant, $F(1, 366) = 175.79$, $p < .001$. Moreover, the target gender effect within each stereotype index was also significant (see Figure 12). Specifically, participants had an easier time associating men with scientific traits ($F(1, 366) = 30.43$, $p < .001$) and with cold traits ($F(1, 366) = 42.75$, $p < .001$) relative to women; and an easier time associating women with warm traits ($F(1, 366) = 107.86$, $p < .001$) and with unscientific traits ($F(1, 366) = 26.11$, $p < .001$) relative to men. Thus, implicit gendered trait stereotypes were present for each stereotype index.

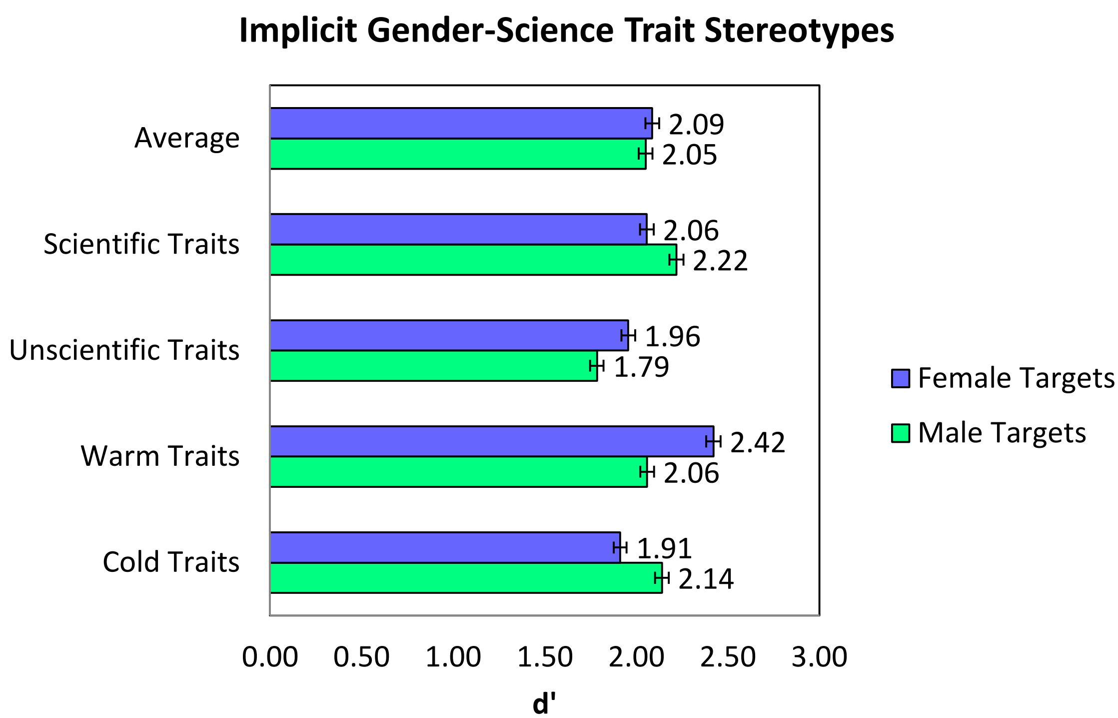


Figure 12. Implicit gender-science stereotypes by target gender.

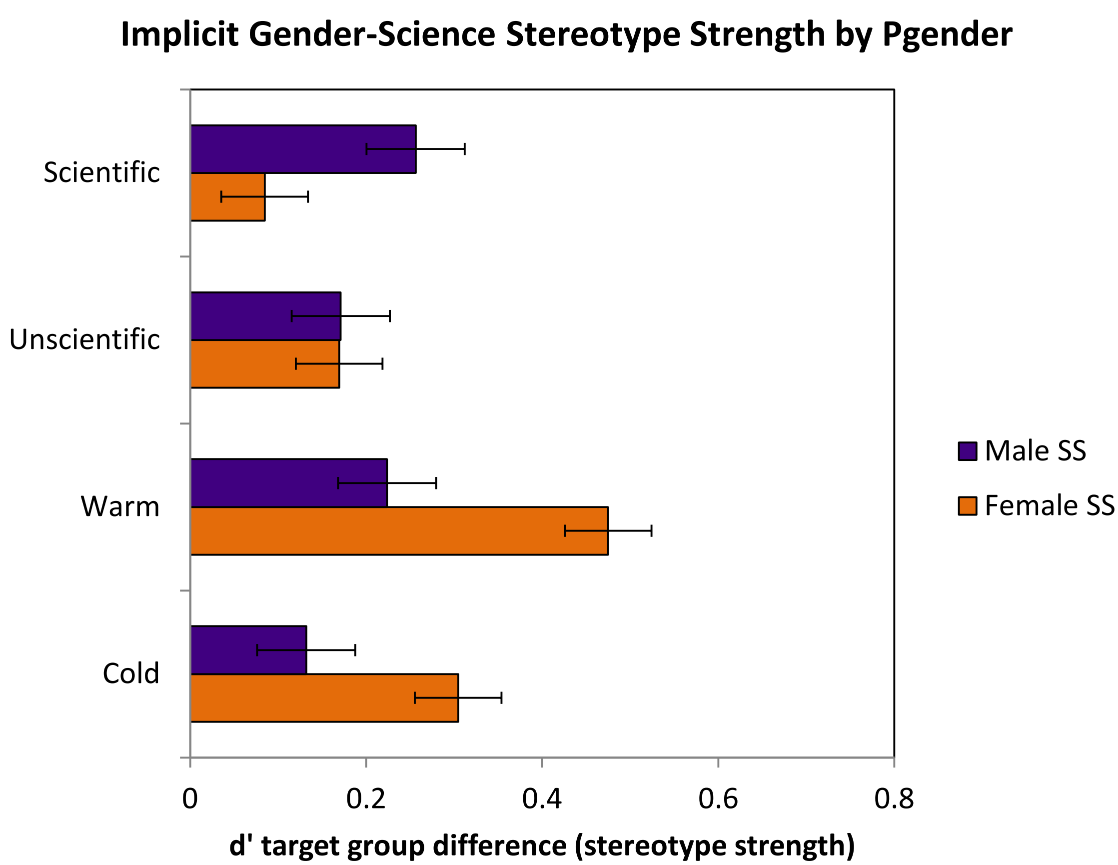
Results from the ANOVA are presented in Table 14. Again, there were other significant lower order effects present, but all were moderated by the primary 3-way interaction of interest.

Effect	F	df	p
Target Gender	3.78	1, 366	.053
Domain	54.80*	1, 366	<.001
Valence	193.90*	1, 366	<.001
Target Gender × Domain	3.76	1, 366	.053
Target Gender × Valence	10.59	1, 366	.001
Domain × Valence	3.77	1, 366	.053
Target Gender × Domain × Valence	176.79*	1, 366	<.001
PGender Effects			
PGender	.54	1, 366	.46
Domain × PGender	7.70*	1, 366	.006
Valence × PGender	6.57*	1, 366	.01
Target Gender × PGender	5.22*	1, 366	.02
Target Gender × Domain × PGender	.16	1, 366	.69
Target Gender × Valence × PGender	28.34*	1, 366	<.001
Domain × Valence × PGender	9.64*	1, 366	.002
Target Gender × Domain × Valence × PGender	3.86	1, 366	.050

Table 15. ANOVA output for implicit overall stereotypes. PGender = Participant Gender. *Significant at $p < .05$. Although not presented in the table, Condition was orthogonally contrast coded and was also included as a factor in the analysis.

As seen in Table 15 and Figure 13, overall stereotypes were moderated by participant gender. This indicated that in contrast to explicit stereotypes, females had *stronger* overall implicit stereotypes than males. Looking within each index, females had stronger implicit stereotypic associations for both warm (participant gender effect; $F(1, 366) = 20.61$, $p < .001$) and cold traits (participant gender effect: $F(1, 366) = 7.40$, $p < .01$). That is, female participants associated women with warm more than men and men with cold more than women to a stronger degree than male participants—precisely the opposite of what occurred for explicit ratings. Importantly, stereotypes were still present for both genders: for warm, all $Fs(1,366) \geq 14.99$, $p < .001$, for cold, all Fs$(1, 366) \geq 6.40$, $p \leq .01$ (see Figure 13).

Although females possessed stronger implicit stereotypes on both the warm and cold dimensions, males had stronger implicit stereotypes for scientific traits, $F(1, 366) = 9.19$, $p < .01$. In fact, for female participants, gender stereotypes for science traits were only marginally significant, $F(1, 366)= 3.59$, $p = .06$ (see Figure 13). All other implicit stereotype indices were highly significant within both male and female participants. In sum, implicit gender-science stereotypes existed for each stereotype index, and were generally robust across males and females. That said, the patterns of associations, tended to favor one's in-group – on the positive dimensions (scientific and warm), each gender group expressed stronger implicit associations for the dimension stereotypically paired with their group (i.e., females had stronger warm stereotypes, males had stronger scientific stereotypes). On cold, a negative dimension, males demonstrated weaker stereotypic associations, and females stronger, again demonstrating a pattern that favors the in-group. Unscientific traits showed significant implicit stereotypes (unlike the explicit judgments), and their strength did not depend on participant gender.



G

Figure 13. Implicit stereotype strength by participant gender. Stereotype Strength was defined as the mean rating for the stereotypic minus counter-stereotypic group. Gender Stimuli Condition was partialed out of the means.

Gender Stimuli Condition Effects

Condition was contrast-coded to test the primary contrast of interest, Low Prototypic vs. High Prototypic Gender Stimuli. Generic Names vs. Photos (Low Prototypic and High Prototypic Gender Stimuli) was included as its orthogonal contrast code. Hypothesis 4 maintained that *implicit stereotypes will be moderated by gender prototypical facial appearance: feminine female and masculine male scientists (i.e., high prototypical) should elicit stronger gendered science trait associations than masculine women and feminine men (i.e., low prototypical).* In other words, the Target Gender × Domain × Valence interaction capturing stereotypes should depend on whether photos are low or high in gender prototypicality—the Low Prototypic vs. High Prototypic effect. There was not a strong hypothesis about how the strength of implicit associations to generic names would compare to the Low Prototypic and High Prototypic conditions.

The data did not support the hypothesis that High Prototypic gender stimuli would exacerbate gender-science stereotypes relative to Low Prototypic gender stimuli: the three-way interaction indicating overall stereotype strength (Target Gender × Domain × Valence) did not depend on whether photos were high or low in gender prototypicality (see Table 16 for all results related to Condition). However, overall stereotype strength was qualified by a Condition (Low Prototypic vs. High Prototypic) × Participant Gender interaction (see Table 16). Looking within participant gender revealed that the four-way interaction of interest was indeed significant for female participants, $F(1, 366) = 4.50$, $p = .03$, but not for male participants, $F(1, 366) = .79$, $p = .37$. Looking within each trait index for female participants, the Low Prototypic vs. High Prototypic effect was significant for unscientific stereotypes. Here, High Prototypic photos unexpectedly reduced stereotypes relative to Low Prototypic photos, $F(1, 366) = 7.25$, $p < .01$. As seen in Figure 14, females in the High Prototypic condition actually did not show any evidence of implicit stereotypes regarding unscientific traits, $F(1, 366) = .03$, $p = .87$, whereas they did show stereotypes in the Low Prototypic condition, $F(1, 366) = 13.04$, $p < .001$. Warm stereotypes were also marginally weaker in the High Prototypic than the Low Prototypic condition, $F(1, 366) = 3.55$, $p = .06$, although the simple target effect was still present in both conditions, $Fs(1, 366) \geq 29.83$, $ps < .001$. There were no differences due to condition for scientific or cold traits, *p*s > .54. In sum, for female participants only, High Prototypic photos eliminated implicit stereotypes regarding unscientific traits, and weakened implicit stereotypes regarding warm traits, relative to Low Prototypic photos.

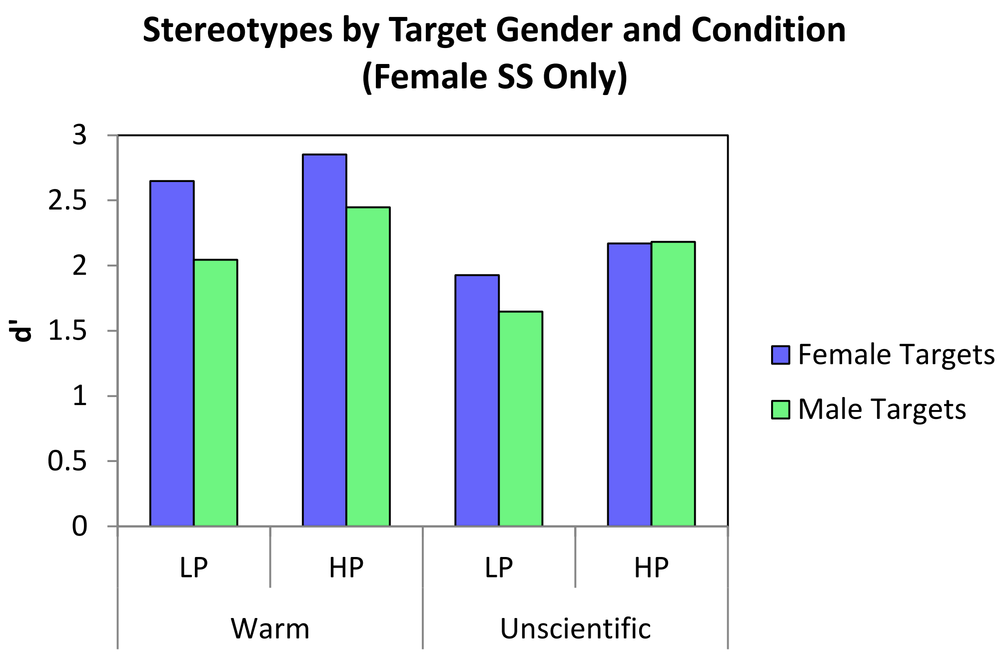


Figure 14. Warm and unscientific stereotypes by target gender and photo condition for female participants only. LP = Low Prototypic Stimuli. HP = High Prototypic Stimuli.

Finally, a four-way interaction showed that the strength of the overall stereotype effect depended on whether participants saw generic gender names vs. photos (i.e., the Target Gender × Domain × Valence × Names vs. Photos interaction; see Table 16). Looking within each stereotype index, the Names vs. Photos effect was only present for warm trait stereotypes, $F(1, 366) = 5.00$, $p = .02$, for all other indices, $Fs(1, 366) \leq .47$, $ps \geq .49$. Specifically, photo stimuli *weakened* associations of women with warm traits relative to men with warm traits compared to generic name stimuli. Importantly, warm trait stereotypes were still robust in the photo conditions, as the simple target effect for the warm stereotype within photo conditions was still substantial, $F(1, 368) = 50.98$, $p < .001$.

In summary, there was minimal evidence that the prototypicality of the gender presented affected implicit stereotypes, and when it did occur (for female participants on the unscientific dimension), it was in the opposed direction as predicted—High Prototypic stimuli did not elicit the unscientific stereotypic, whereas Low Prototypic stimuli did. Because stimuli type did not impact stereotypes in a systematic way, there was no reason to examine gender ideology's ability to strengthen or weaken the effect of target prototypicality on gender bias.

Condition Effects	*F*	*df*	*p*
Names v. Photos	55.53*	1, 366	< .001
LP v. HP	32.10*	1, 366	< .001
TGender × Names v. Photos	.72	1, 366	.39
TGender × LP v. HP	8.21*	1, 366	.004
Domain × Names v. Photos	3.65	1, 366	.057
Domain × LP v. HP	1.19	1, 366	.28
Valence × Names v. Photos	.36	1, 366	.55
Valence × LP v. HP	1.57	1, 366	.21
TGender × Domain × Names v. Photos	.59	1, 366	.44
TGender × Domain × LP v. HP	.03	1, 366	.87
TGender × Valence × Names v. Photos	.60	1, 366	.44
TGender × Valence × LP v. HP	.20	1, 366	.65
Domain × Valence × Names v. Photos	.20	1, 366	.66
Domain × Valence × LP v. HP	.16	1, 366	.69
TGender × Domain × Valence × Names v. Photos	4.18*	1, 366	.04
TGender × Domain × Valence × LP v. HP	.61	1, 366	.43
Participant Gender Effects			
PGender × Names v. Photos	.33	1, 366	.56
PGender × LP v. HP	.70	1, 366	.40
PGender × TGender × Names v. Photos	2.24	1, 366	.13
PGender × TGender × LP v. HP	.10	1, 366	.75
PGender × Domain × Names v. Photos	4.76*	1, 366	.03
PGender × Domain × LP v. HP	2.07	1, 366	.15
PGender × Valence × Names v. Photos	1.89	1, 366	.17
PGender × Valence × LP v. HP	1.26	1, 366	.26
PGender × TGender × Domain × Names v. Photos	.69	1, 366	.41
PGender × TGender × Domain × LP v. HP	2.17	1, 366	.14
PGender × TGender × Valence × Names v. Photos	6.48*	1, 366	.01
PGender × TGender × Valence × LP v. HP	.38	1, 366	.54
PGender × Domain × Valence × Names v. Photos	4.58*	1, 366	.03
PGender × Domain × Valence × LP v. HP	2.35	1, 366	.12
PGender × TGender × Domain × Valence × Names v. Photos	.61	1, 366	.44
PGender × TGender × Domain × Valence × LP v. HP	4.50*	1, 366	.03

Table 16. Condition effects in ANOVA output for implicit overall stereotypes. *Significant at $p < .05$. TGender = Target Gender. PGender = Participant Gender. LP = Low Prototypic Stimuli. HP = High Prototypic Stimuli.

Overall sensitivity was also examined as a function of Target Gender, Condition, Participant Gender, and their interactions. As seen in Figure 15, on average across blocks, performance on the GNAT was superior (i.e., d' was higher or sensitivity was greater) for those in the photograph than the name condition, $F(1, 366) = 54.91$, $p < .001$, and greater for stereotypic photos than for counter-stereotypic photos, $F(1, 366) = 32.52$, $p < .001$.[12] There were no other significant effects on overall sensitivity. Thus, participants had an easier time completing the GNAT when presented with High Prototypic photos, followed by Low Prototypic photos, and finally, with generic gender names.

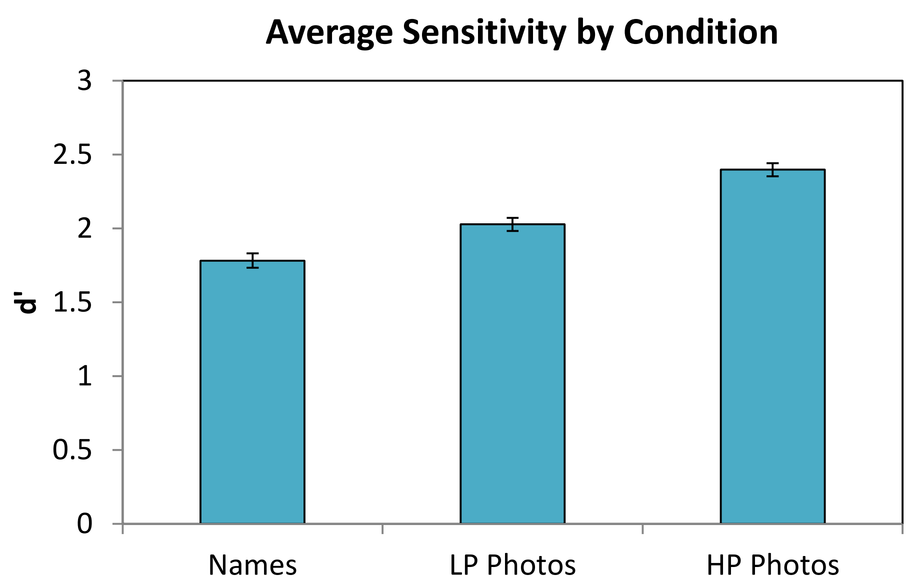


Figure 15. Average sensitivity on the GNAT by gender stimuli. LP = Low Prototypic Stimuli. HP = High Prototypic Stimuli.

Gender Ideology

I next examined the extent to which the four gender ideologies, gender blindness, gender awareness, segregationism, and assimilationism, affected gender-science stereotypes. The means and standard deviations for each gender ideology are presented in Table 17.

[12] The higher d's may have been due to more easily processing, for example, the gender prototypical faces, without affecting responses to the trait words. To examine this, d' values based only on responses to trait stimuli were calculated. Basing the d's just on responses to the trait items, d' was still greater in the photo conditions relative to the generic name condition, $F(1, 366) = 4.61$, $p = .03$, and in the High Prototypic Condition compared to the Low Prototypic Condition, $F(1, 366) = 17.09$, $p < .001$. Thus performance on the entire GNAT block was enhanced in the photo condition relative to names, and in the High Prototypic condition relative to Low Prototypic condition, suggesting perhaps greater fluency of processing in the prototypical face condition.

Ideology	*M (SD)*		Alpha	N Items
	Female SS	Male SS		
Gender Aware	5.60 (.70)	5.60 (.72)	.71	5
Gender Blind	5.71 (.68)	5.56 (.73)*	.45	4
Segregationism	3.09 (.80)	3.61 (.92)*	.64	5
Assimilationism	3.12 (1.12)	3.33 (1.11)+	.78	4

Table 17. Gender Ideology descriptive statistics. *Males differed from females, $p < .05$. +Males differed from females, $p = .08$. $N = 359$.

Endorsement of the ideologies was submitted to a 2 (Valence: Positive vs. Negative) × 2 (Differentiation: high vs. Low) × 2 (Participant Gender) ANOVA. This revealed that participants endorsed the positive ideologies significantly more than the negative ideologies, $F(1, 357) = 1674.45$, $p < .001$, and this was especially the case for female participants, $F(1, 357) = 14.93$, $p < .001$. There was not a main effect of differentiation, meaning participants similarly endorsed acknowledging vs. minimizing gender differences, $p = .23$. A Differentiation × Participant Gender interaction, however, $F(1, 357) = 10.10$, $p < .01$, showed that female participants generally preferred ignoring gender differences to acknowledging them, relative to men. Looking at the means, this was largely driven by females' far lower endorsement of Segregation. Finally, a significant Valence × Differentiation interaction showed that the difference between gender blindness and gender awareness was smaller than that between segregationism and assimilationism, $F(1, 357) = 5.18$, $p = .02$, which did not depend on participant gender, $p = .25$.

Moderation of Explicit Gender-science Stereotypes.

Hypothesis 5 maintained that *perceiver gender ideologies, or beliefs about how to approach and deal with gender differences, would moderate explicit gender-science stereotypes.* To test this hypothesis, each trait index was regressed onto participant gender, each gender ideology (mean-centered), and each ideology's interaction with participant gender. Importantly, this allowed an examination of each ideology's impact on stereotyping over and above the other ideologies. Condition was not included because all measures were from the pre-screen and therefore the experimental condition was irrelevant. Unless otherwise noted, the same results described in the previous ANOVA (without gender ideologies) persisted when the gender ideologies were included as a factor. No Participant Gender × Ideology interactions emerged, *F*s(1, 325) ≤ 3.03, *p*s ≥ .08, and so they are not discussed.

	Dependent Variables			
Regressed onto…	Scientific Stereotypes (Men – Women)	Unscientific Stereotypes (Women – Men)	Warm Stereotypes (Women – Men)	Cold Stereotypes (Men – Women)
Blindness	**-.11***	**-.16****	-.05	-.08
Awareness	.02	.03	.06	.03
Assimilationism	**.27****	**.27****	.05	.02
Segregationism	-.06	-.05	-.10	-.05
PGender	-.01	**.12***	**-.26****	**-.27****

Table 18. Standardized betas for explicit stereotype indices as a function of Gender Ideologies. Participant gender was coded as +1 = Female, -1 = Male. *$p < .05$. **$p < .01$. Participant gender was included as a factor.

As can be seen in Table 18, the more that a participant endorsed a gender blind ideology—believing that we should look beyond gender differences and treat people as individuals—the weaker his or her gender stereotypes with respect to both scientific and unscientific traits. On the other hand, those who endorsed assimilationism more—the view that women should become like men in the workplace if they want to succeed—expressed stronger scientific and unscientific stereotypes. There were no other significant effects of gender ideology on stereotypes. Notably, warm and cold stereotypes were unaffected by the gender ideologies.

Moderation of Implicit Gender-science Stereotypes.

Hypothesis 5 also maintained that *gender ideologies would also moderate implicit stereotypes, although potentially to a lesser extent than explicit stereotypes.* To examine this hypothesis, overall implicit stereotypes, as well as each individual stereotype index, were examined as a function of the gender ideologies in a series of 2 (Target Gender) × 2 (Trait Valence) × 2 (Trait Domain) × 2 (Participant Gender) × 3 (Condition) mixed ANOVAs that included each gender ideology as a mean-centered predictor, as well as each ideology's interaction with participant gender. As before, participant gender and condition were orthogonally contrasted coded.

Overall implicit stereotypes did not depend on gender ideology, *p*s > .38, nor did any individual stereotype index, all *F*s(1, 345) > 3.01, *p*s > .08. However, a single marginal Participant Gender × Assimilationism interaction emerged, $F(1, 345) = 3.68$, $p = .056$. Within males, assimilationism was marginally and positively related to stronger overall stereotypes, $F(1, 345) = 3.11$, $p = .08$; whereas within females, assimilationism was unrelated to stereotypes, $p = .37$. Looking at each stereotype index separately for males, assimilationism was related to stronger unscientific stereotypes, $F(1, 345) = 4.39$, $p = .04$. No other stereotype indices were moderated by assimilationism for males.

The Relationship of Implicit And Explicit Stereotypes

Finally, I explored the relationship between implicit and explicit stereotypes, although there was not a strong hypothesis regarding the extent to which they would be related (Hypothesis 6). To do so, each implicit and explicit stereotype index was correlated with each another, partialing out participant gender and condition. As can be seen in Table 19, implicit and explicit gender-science stereotypes were completely uncorrelated with one another. Possible explanations for the lack of relationship are described in the discussion section.

		Implicit Stereotype				
		Overall	Scientific	Un-scientific	Warm	Cold
Explicit Stereotype	Overall	**0.08**	0.07	0.11	0.03	0.04
	Scientific	-0.05	**0.08**	0.06	0.04	-0.08
	Unscientific	0.05	0.05	**-0.05**	-0.04	-0.06
	Warm	-0.05	0.03	0.08	**0.00**	-0.02
	Cold	-0.13	0.00	0.16	0.05	**0.06**
	Personal Ster	-0.06	-0.04	0.09	0.01	0.05
	Societal Ster	-0.08	0.05	0.08	-0.03	0.08

Table 19. Implicit and explicit (pre-screen) stereotype index correlations. Participant Gender and Condition are partialed.

Ancillary Analyses

Additional analyses examined 1) lab session explicit gender-science stereotypes; 2) effect sizes across the various forms of stereotypes; 3) the measurement properties of the gender ideologies and the gender-science stereotypes using confirmatory factor analysis (CFA) and tests of invariance to ensure that the items fit similarly for male and female participants; 4) test re-test reliability from the pre-screen to the lab session explicit measures; and finally 5) implicit evaluation results from the GNAT.

Laboratory Explicit Gender-science stereotypes

Given the lack of a relationship between implicit and explicit traits, as well as the failure to detect a stereotypic difference between men and women regarding explicit *unscientific* stereotypes at the pre-screen, explicit stereotype measures assessed immediately after the GNAT were examined in order to see whether these measures a) revealed explicit unscientific stereotypes or b) showed greater correspondence to the implicit task. Thus, explicit stereotypes were analyzed again using lab session measures instead of pre-screen measures, and using only the traits that appeared in the GNAT to represent unscientific (i.e., ditzy, careless, forgetful, gullible, naïve).

The three-way interaction indicating stereotypes was again strongly significant, $F(1, 360) = 413.65$, $p < .001$. Looking at each stereotype index, men were again judged as more scientific than women, $F(1, 361) = 114.56$, $p < .001$ (although female participants endorsed this less, $F(1, 361) = 22.90$, $p < .001$) and as colder than women, $F(1, 362) = 81.16$, $p < .001$ (although female participants endorsed this less, $F(1, 362) = 20.30$, $p < .001$). On the other hand, women were judged as warmer than men, $F(1, 361) = 692.81$, $p < .001$ (no participant gender difference), *and* as more unscientific than men, $F(1, 361) = 22.96$, $p < .001$ (no participant gender difference). These participant gender differences were also somewhat different from the pre-screen, where females showed no difference on scientific and unscientific traits, but had weaker warm and cold stereotypes.

Thus, explicit trait stereotype measures from the lab largely replicated those from the pre-screen, except that in the lab session, women were in fact seen as significantly more likely to possess *unscientific* traits than men. As one further test, I examined unscientific stereotypes using only the traits selected for the pre-screen (i.e., incompetent, imprecise, gullible, naïve). Using these same traits, unscientific stereotypes still emerged in the lab session, $F(1, 362) = 55.54$, $p < .001$.

However, even when using the *exact* same traits on every index for both explicit judgments and in the GNAT, there was still no relationship between explicit and implicit gender-science stereotypes, *r*s(1, 366) < .04, *p*s > .43 (see Table 20 for the correlations between all stereotype indices including pre-screen, laboratory, and the GNAT). This table shows that implicit associations were not related to explicit judgments in either the pre-screen or the lab session. The single exception is an unanticipated relationship where pre-screen cold stereotypes were related to stronger implicit unscientific stereotypes. In contrast to the lack of correlations with implicit attitudes, pre-screen stereotypes were significantly related to their matching lab session stereotype (with *r*s between .28 to .56, see Table 20).

Source	Stereotype Index	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
Pre-Screen	1. Scientific	1.00										
	2. Unscientific	0.21*	1.00									
	3. Warm	0.11*	0.07	1.00								
	4. Cold	0.15*	-0.16*	0.09	1.00							
GNAT	5. Scientific	**0.08**	0.05	0.03	0.00	1.00						
	6. Unscientific	0.06	**-0.05**	0.08	0.16*	-0.03	1.00					
	7. Warm	0.04	-0.04	**0.00**	0.05	-0.01	0.04	1.00				
	8. Cold	-0.08	-0.06	-0.02	**0.06**	0.00	0.11	0.05	1.00			
Laboratory	9. Scientific	**0.28***	0.20*	0.03	0.07	**-0.03**	0.11	0.03	-0.04	1.00		
	10. Unscientific	0.20*	**0.32***	0.01	-0.05	-0.02	**0.00**	0.04	-0.07	0.43*	1.00	
	11. Warm	0.06	0.07	**0.56***	0.07	-0.01	0.06	**0.03**	-0.10	0.23*	0.19*	1.00
	12. Cold	0.13*	0.04	0.09*	**0.29***	0.01	0.13	0.08	**0.03**	0.39*	0.10	0.26*

Table 20. Correlations of all stereotype indices. Participant Gender and Condition are partialed. N = 330. *Indicates significant at $p < .05$. Unscientific was defined using the GNAT traits.

In sum, there is evidence that people possessed strong explicit and implicit gender-science stereotypes, viewing women as warmer and more unscientific than men, and men as colder and more scientific than women. There was little evidence that the gender stimuli affected implicit gendered trait stereotypes, except that photos minimized bias relative to generic names. Contrary to hypotheses, whether photos were low or high on gender prototypicality had minimal impact on gender stereotypes. If anything, for female participants, High Prototypic photos actually resulted in weaker implicit stereotypes than Low Prototypic, and particularly for the unscientific dimension.

There was some evidence that gender ideologies, and in particular gender blindness and assimilationism, played a part in shaping explicit scientific and unscientific stereotypes, but not warm and cold stereotypes. Gender blindness minimized these stereotypes, and assimilationism exacerbated them. There was no evidence that gender ideology affected implicit stereotypes, however, and there was no relationship between implicit and explicit trait measures. Given the similar pattern of mean-level stereotypes across implicit and explicit domains, this was somewhat surprising.

A Comparison of Effect Sizes

Effect sizes for the various stereotyping effects were compared across explicit and implicit measures. Condition was controlled for implicit and lab session measures (see Table 21). Partial eta-squared represents the proportion of variance in an outcome that is due to the predictor, controlling for other predictors in the model (Cohen, 1973). Across these three different measures, effect sizes for each stereotype dimension were strikingly similar. Across each measure, warmth domain stereotypes (i.e., warm and cold stereotype indices) showed larger effect sizes than scientific domain stereotypes (i.e., scientific and unscientific).

	Career Stereotypes (Scientist vs. Teacher)	Explicit Gender Stereotypes (Men vs. Women)		Implicit Gender Stereotypes (Men vs. Women)
Stereotype	Pre-Test	Pre-Screen	Lab	GNAT
Overall	.90	.59	0.52	.33
Scientific	.83	.16	.24	.08
Unscientific	.16	.00	.03	.07
Warm	.86	.69	.66	.23
Cold	.80	.18	.18	.11

Table 21. Effect Sizes (η^2_p) for stereotypes in Pre-Test, Pre-Screen, Lab, and GNAT. All partial eta-squares are from models that included participant gender. For Lab and GNAT, models included condition and its interaction with participant gender. η^2_p of .01 constitutes a small effect, .06 a medium effect, and .14 a large effect.

Other Relationships of Interest

I also explored the relationship of implicit and explicit gender-science stereotypes to other measures of interest that were collected during the laboratory session. In the laboratory session, a short version of the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996) and the complete Internal vs. External Motivation to Response without Sexism scale (IMS/EMS; Klonis, et al., 2005) were included. These were presented on separate pages with items in each section presented in a uniquely randomized order for each participant (see Appendix F for all items). The ASI posits two different forms of sexism that result in ambivalence toward women: benevolent sexism, a subjectively positive (for sexist men) orientation that relates to admiring and cherishing traditional women (e.g., every man ought to have a woman whom he adores), and hostile sexism, which captures sexist antipathy for women, particularly nontraditional women (e.g., women seek to gain power by getting control over men). Whereas hostile sexism is related to negative stereotypes and attitudes about women, benevolent sexism is related to positive stereotypes and attitudes about women. Four items each assessed hostile and benevolent sexism, with ratings from 1 (Strongly Disagree) to 6 (Strongly Agree). The 10-item Internal and External Motivation to Respond without Sexism scale (IMS/EMS), was also included. This scale assesses the degree to which people are internally versus externally motivated to respond in a non-sexist manner (5 items each, e.g., "*I am personally motivated by my beliefs to be nonsexist toward women*", "*I try to act in nonsexist ways because of pressure from others*", respectively), from 1 (Strongly Disagree) to 7 (Strongly Agree).

The thermometer measurements included in the lab assessed warmth towards men and women as well as subgroups of men and women that were intended to represent traditional gender roles and non-traditional gender roles. For traditional roles, these consisted of male politician, male scientist, female nurse, and stay at home mom (α = .55). For non-traditional roles, they consisted of female politician, female scientist, male nurse, and stay at home dad (α = .63). Descriptive statistics for all measures examined are presented in Table 22.

Index	Index/Variable	N Responses	N Items	M (SD)	Alpha
Thermometer (0-100)	Thermometer Women	348	1	71.46 (19.39)[b]	--
	Thermometer Men	348	1	67.34 (20.28)[a]	--
	Therm Women (Traditional)*	372	2	77.88 (14.52)	.39
	Therm Men (Traditional)*	372	2	58.06 (19.64)	.44
	Therm Women (Non-Trad)*	372	2	59.08 (19.49)[a]	.41
	Therm Men (Non-Trad)*	372	2	63.32 (20.72)[a]	.49
Gender STEM Stereotypes (1-7)	Personal	359	1	3.32 (1.54)[b]	--
	Societal	359	1	4.72 (1.50)	--
Ambivalent Sexism (1-6)	Hostile Sexism*	371	5	3.51 (.90)[b]	.82
	Benevolent Sexism*	370	5	3.94 (.81)[b]	.65
IMS/EMS (1-7)	Internal Motivation*	371	5	5.06 (1.04)	.81
	External Motivation*	371	5	3.82 (1.11)	.78

Table 22. Descriptive statistics for all other indices collected. *Indicates from laboratory session. [a]Female SS > Male SS at $p < .05$; [b]Male SS > Female SS at $p < .05$.

Correlations of all indices (from the pre-screen where available, from lab session otherwise), partialing out participant gender, are presented in Table 23. Some interesting relationships emerged. Not surprisingly, scientific trait stereotypes were positively related to the following: personal belief that men are better at math and science than are women (i.e., personal STEM); the perception that society also believes men are better at math and science than women (i.e., societal STEM stereotype); and hostile sexism. Scientific stereotypes were also negatively related to internal motivation to respond without sexism.

Unscientific trait stereotypes were positively related to personal STEM stereotypes and hostile sexism, and negatively related to internal motivation to respond without sexism. Warm and cold trait stereotypes were unrelated to personal and societal STEM stereotypes. Warm trait stereotypes, however, were positively related to both hostile and benevolent sexism.

Endorsement of the personal gender STEM stereotype was moderately related to societal gender STEM stereotypes. Moreover, it was positively related to both hostile and benevolent sexism, and negatively with internal motivation to respond without sexism. Endorsement of the societal gender-STEM stereotype was not as meaningful, and only related to hostile sexism. Although an evaluative preference for women was unrelated to other indices, greater warmth towards *traditional* rather than *non-traditional* targets was positively related to hostile and benevolent sexism, negatively related to internal motivation to respond without sexism, and positively related to external motivation to respond without sexism.

	Explicit Trait Stereotypes				Gender-STEM Stereotypes		Sexisms and Motivation				Therm
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Scientific Stereotype	1.00										
2. Unscientific Stereotype	**0.21**	1.00									
3. Warm Stereotype	**0.11**	0.07	1.00								
4. Cold Stereotype	**0.14**	**-0.16**	0.08	1.00							
5. Personal STEM Ster	**0.31**	**0.11**	0.03	0.03	1.00						
6. Societal STEM Ster	**0.14**	0.05	0.01	0.09	**0.44**	1.00					
7. Hostile Sexism*	**0.17**	**0.18**	**0.12**	-0.03	**0.36**	**0.14**	1.00				
8. Benevolent Sexism*	0.02	0.00	**0.16**	-0.02	**0.22**	0.09	**0.39**	1.00			
9. Internal Motivation*	**-0.25**	**-0.20**	0.04	-0.01	**-0.31**	-0.07	**-0.34**	-0.05	1.00		
10. External Motivation*	0.01	0.07	0.01	-0.02	0.04	0.04	**0.19**	**0.14**	**0.13**	1.00	
11. Therm Index	-0.05	-0.06	0.01	0.02	-0.03	0.07	-0.05	0.06	0.09	0.03	1.00
12. Therm Trad v. Non-Trad*	0.10	0.02	0.08	-0.07	0.07	0.05	**0.25**	**0.21**	**-0.11**	**0.16**	**0.13**

Table 23. Correlations of explicit stereotype indices with other explicit measures. Therm Index = women – men. Therm Trad v. Non-Trad = all traditional targets – all non-traditional targets. Participant Gender and Condition were partialed. Complete N = 328. *Measured in lab session; otherwise from pre-screen. Correlations ≥ .11 (bolded) were significant at $p < .05$ level.

Table 24 examines these same relationships but rather than presenting relative indices, it presents the simple trait stereotype effects of target gender. Not surprisingly, the table showed that ASI, internal motivation to respond without sexism/external motivation to respond without sexism, and even thermometer ratings showed more numerous meaningful relationships for target women than men, indicating that these gender-related ideologies beliefs tend to carry more weight for judgments of women than men.

	Explicit Judgments							
	Target Women				Target Men			
	1.	2.	3.	4.	1.	2.	3.	4.
1. Percent Scientific	1.00				1.00			
2. Percent Unscientific	0.07	1.00			**0.14**	1.00		
3. Percent Warm	**0.43**	0.04	1.00		**0.41**	0.10	1.00	
4. Percent Cold	**0.33**	**0.58**	**0.16**	1.00	**0.33**	**0.57**	0.09	1.00
Blind	**0.18**	-0.1	0.09	0.01	0.06	0.04	0.10	-0.04
Aware	0.05	0	0.1	0.04	0.07	-0.04	0.03	0.05
Assimilation	**-0.14**	**0.26**	0.07	**0.13**	**0.12**	0.07	0.04	**0.14**
Segregation	-0.07	**0.22**	0.04	**0.17**	0.03	**0.15**	0.09	**0.14**
Personal STEM Ster	**-0.18**	0.06	0.04	0.07	**0.11**	-0.02	0.00	0.09
Societal STEM Ster	-0.06	0.02	0.07	0.07	0.06	-0.02	0.04	**0.14**
Hostile Sexism*	**-0.15**	**0.23**	0.06	**0.19**	0.01	0.10	-0.07	**0.16**
Benevolent Sexism*	0.00	0.05	**0.16**	0.09	0.03	0.06	-0.03	0.07
Internal Motivation*	**0.19**	**-0.20**	0.13	-0.07	-0.05	-0.06	0.06	-0.07
External Motivation*	0.03	0.08	**0.11**	0.07	0.04	0.02	0.08	0.05
Therm General	**0.25**	-0.09	**0.25**	-0.04	0.10	**-0.11**	**0.13**	-0.07
Therm Traditional*	0.05	-0.05	0.08	**-0.13**	0.02	-0.06	0.06	-0.07
Therm Non-Trad*	**0.14**	**-0.19**	0.04	**-0.16**	0.02	**-0.15**	0.10	-0.07

Table 24. Explicit trait stereotype correlations separated by target gender. Participant Gender and Condition are partialed. Complete N = 333. *Measure was from lab session. Otherwise from pre-screen. Correlations ≥ .11 (bolded) were significant at $p < .05$ level.

Finally, Table 25 depicts relationships between gender ideologies and other indices. Not surprisingly, gender blindness was related to weaker personal STEM stereotypes and hostile sexism, whereas assimilationism and segregationism were related to stronger personal STEM stereotypes and hostile sexism. Both Assimilation and Segregation were also related to stronger perceptions of societal endorsement of gender STEM stereotypes. Benevolent sexism was positively related to all ideologies except gender blindness , which showed no relationship. internal motivation to respond without sexism was positively related to gender blindness and gender awareness, but negatively related to assimilationism and segregationism. On the other hand, external motivation to respond without sexism was positively related to assimilationism and segregationism, but unrelated to gender awareness and gender blindness. Gender blindness was also uniquely related to positively evaluating non-traditional men and women, where assimilationism and segregationism were negatively related to more negative evaluations of non-traditional women and men relative to their more traditional counter-parts.

		Blind	**Aware**	**Assim**	**Seg**
Gender-STEM Stereotypes	Personal STEM Ster	**-0.19**	-0.07	**0.45**	**0.40**
	Societal STEM Ster	-0.05	-0.02	**0.16**	**0.14**
Sexism	Hostile Sexism*	**-0.11**	-0.01	**0.39**	**0.29**
	Benevolent Sexism*	-0.04	**0.14**	**0.22**	**0.20**
	Internal Motivation*	**0.29**	**0.16**	**-0.43**	**-0.29**
	External Motivation*	-0.10	-0.10	**0.13**	**0.15**
Thermometer	Therm Index	0.05	0.00	0.04	-0.04
	Therm Trad v. Non-Trad	**-0.17**	-0.02	**0.27**	**0.16**

Table 25. Correlations of gender ideologies with other indices. Therm Index = women – men. Therm Trad v. Non-Trad = all traditional targets – all non-traditional targets. Participant Gender and condition were partialed. Correlations ≥ absolute value of .11 are significant at $p < .05$ (in bold). *Measured in laboratory session, otherwise from pre-screen.

As can be seen in Table 26, implicit measures were not related to any explicit measures of interest. Given the many and robust relationships between explicit correlations, it was somewhat surprising that implicit stereotypes did not relate more strongly to other measures.

	Implicit Trait Stereotypes			
	1.	2.	3.	4.
1. Scientific Stereotype	1.00			
2. Unscientific Stereotype	-0.03	1.00		
3. Warm Stereotype	-0.03	0.04	1.00	
4. Cold Stereotype	0.00	0.11	0.05	1.00
5. Personal STEM Ster	-0.03	0.08	0.00	0.05
6. Societal STEM Ster	0.02	0.07	-0.02	0.06
7. Hostile Sexism*	0.10	0.07	-0.04	-0.06
8. Benevolent Sexism*	-0.05	-0.03	0.01	-0.08
9. Internal Motivation*	0.09	-0.07	0.00	-0.04
10. External Motivation*	0.05	-0.04	0.06	-0.02
11. Therm Index	0.00	0.01	0.00	0.01
12. Therm Trad v. Non-Trad	0.01	-0.01	0.02	0.01

Table 26. Correlations of implicit stereotypes with other explicit measures. Therm Index = women – men. Therm Trad v. Non-Trad = all traditional targets – all non-traditional targets. Participant Gender and Condition were partialed. N = 344. *Indicates from Qualtrics, otherwise from pre-screen.

In sum, there were robust gender-science stereotypes for both explicit and implicit traits that were endorsed by both male and female participants. Explicit trait stereotypes were somewhat moderated by gender ideologies, but implicit trait stereotypes were not. In fact, there was no relationship between implicit and explicit stereotypes. Condition was largely inconsequential for implicit gender stereotypes, although both sets of photos—stereotypic and counter-stereotypic—produced weaker implicit stereotypes regarding warm traits compared to generic names. Moreover, photos improved overall performance on the GNAT relative to generic names, and especially high prototypic photos. One small interaction with participant gender indicated that for female participants, stereotypic photos actually produced somewhat weaker implicit stereotypes than counter-stereotypic photos for unscientific and warm stereotypes, although simple effects were not significant.

Measurement

Gender Ideologies.

The gender ideologies were examined in a CFA in which each latent factor variance was set to 1 to identify and standardize the model. Full information maximum likelihood estimation was employed to account for some missing responses across items. Model fit was acceptable, χ^2(*N* = 359, df = 129) = 225.68, *p* < .001, *RMSEA* = .046 [95% CI: .03:.05], *SRMR* = .06, *CFI* = .92, *GFI* = .93 (see Figure 16).

I next examined whether model fit was invariant across men and women. In other words, did male and female participants interpret the items in the same way, and did the model fit equivalently for male and female participants? To examine this question, items with the greatest factor loading in the standardized *CFA* were set to 1. A completely unconstrained model was examined for each group, and subsequent constrains were added that set 1) item loadings to be equal across men and women, 2) factor variances to be equal across men and women, and 3) factor covariances to be equal across men and women. Because there was not a significant increase in χ^2 as the model was constrained, and because *CFI* and *RMSEA* did not change more than .01, there was evidence of invariance—the model fit men and women equally well, despite mean differences in ideology endorsement (see Table 27).

Model	**χ^2(df)**	**Change χ^2**	***CFI***	***RMSEA***
Unconstrained	376.22(258)	--	.90	.05
Add fixed loadings	396.50(272)	20.28(14)	.90	.05
Add fixed factor variances	402.38(276)	5.88(4)	.90	.05
Add fixed latent factor covariances	404.83(282)	2.45(6)	.90	.05

Table 27. Gender Ideology invariance in Pre-Screen. All *p* values are non-significant for the chi-square change.

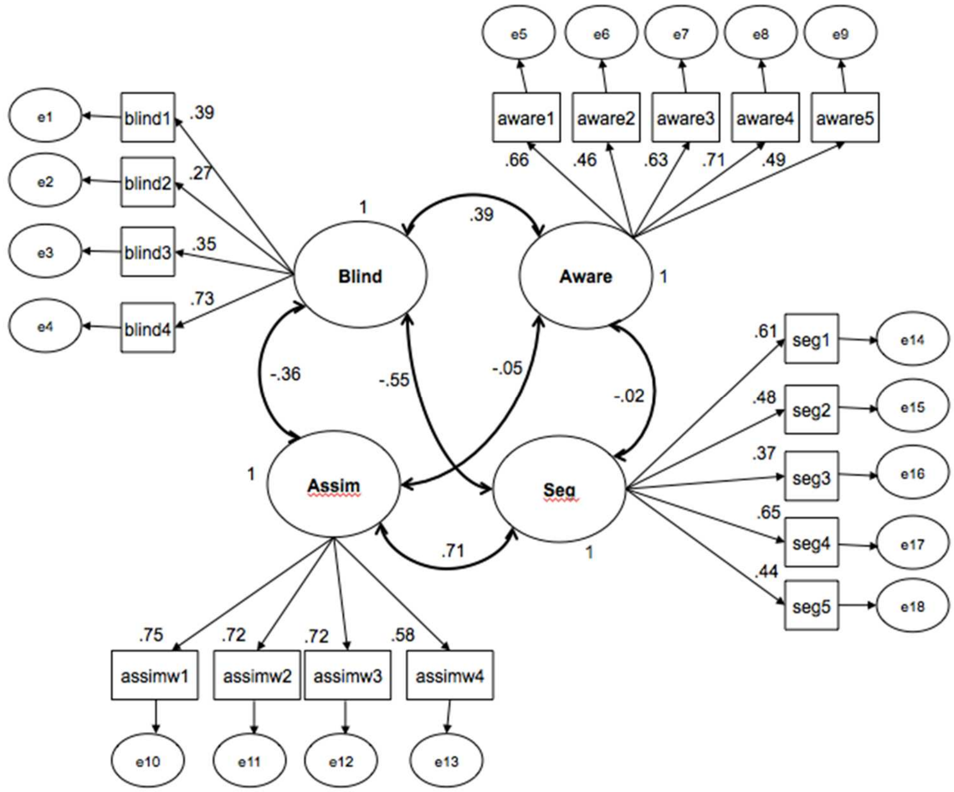


Figure 16. Standardized CFA for Gender Ideology.

Trait Stereotypes.

Next, a confirmatory factor was conducted separately for trait ratings of females and trait ratings of males, in which each factor variance was set to 1. For females targets, model fit was acceptable, $\chi^2(98) = 257.89$, *RMSEA* = .07 [95% *CI*, .06: .08], *CFI* = .89, *GFI* = .90, *SRMR* = .087 (Figure 17); it was also acceptable for male targets, $\chi^2(98) = 202.25$, *RMSEA* = .06 [95% CI, .05:.07], *CFI* = .93, *GFI* = .92, *SRMR* = .07 (Figure 18). All item loadings were significant onto their factors, *t*'s > 5.99 for female targets (the lowest loading was for the trait "critical"), and *t*'s > 7.50 for male targets (again for the trait "critical").

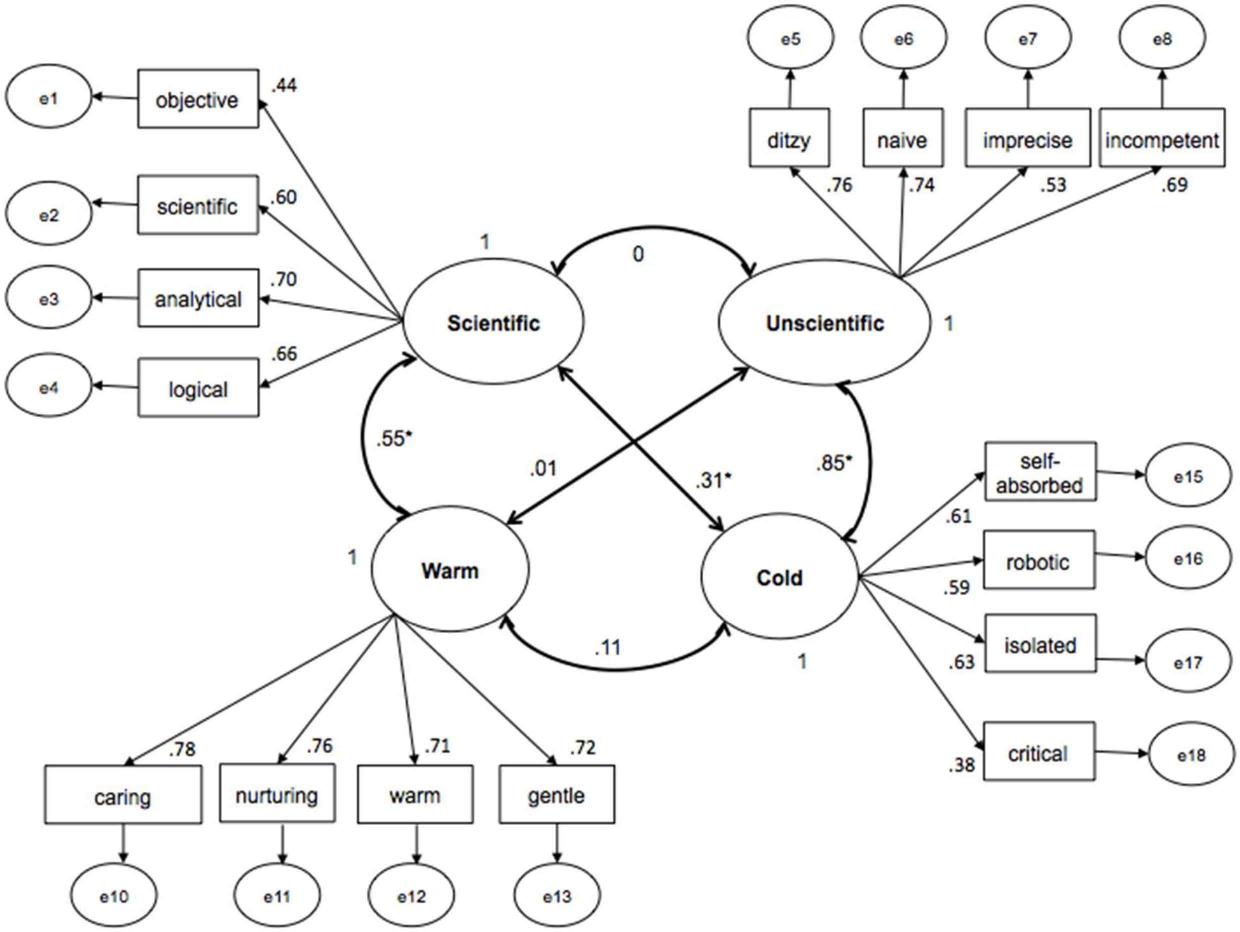


Figure 17. Standardized CFA for traits for female targets.

For both male and female targets, *unscientific* traits and *cold* traits were very highly correlated (r = .85), indicating that valence strongly drove responses to these traits, and suggesting that they should potentially be collapsed into a single “negativity” factor. I examined an alternative model in which all unscientific and cold traits loaded onto one factor; fit was significantly worse for these three-factor models for both female and male targets as evidenced by higher AICs and greater chi-squares in the 3-factor model: for females, 4-factor AIC = 333.89; 3-factor AIC = 356.29, $\Delta\chi^2(3)$= 31.12, p < .001; for males, 4-factor AIC = 278.25, 3-factor AIC = 303.68, $\Delta\chi^2(3)$= 28.40, p < .001. Despite the shared negative valence creating a high correlation between unscientific and cold traits, all four factors were required to adequately capture variation in the trait ratings.

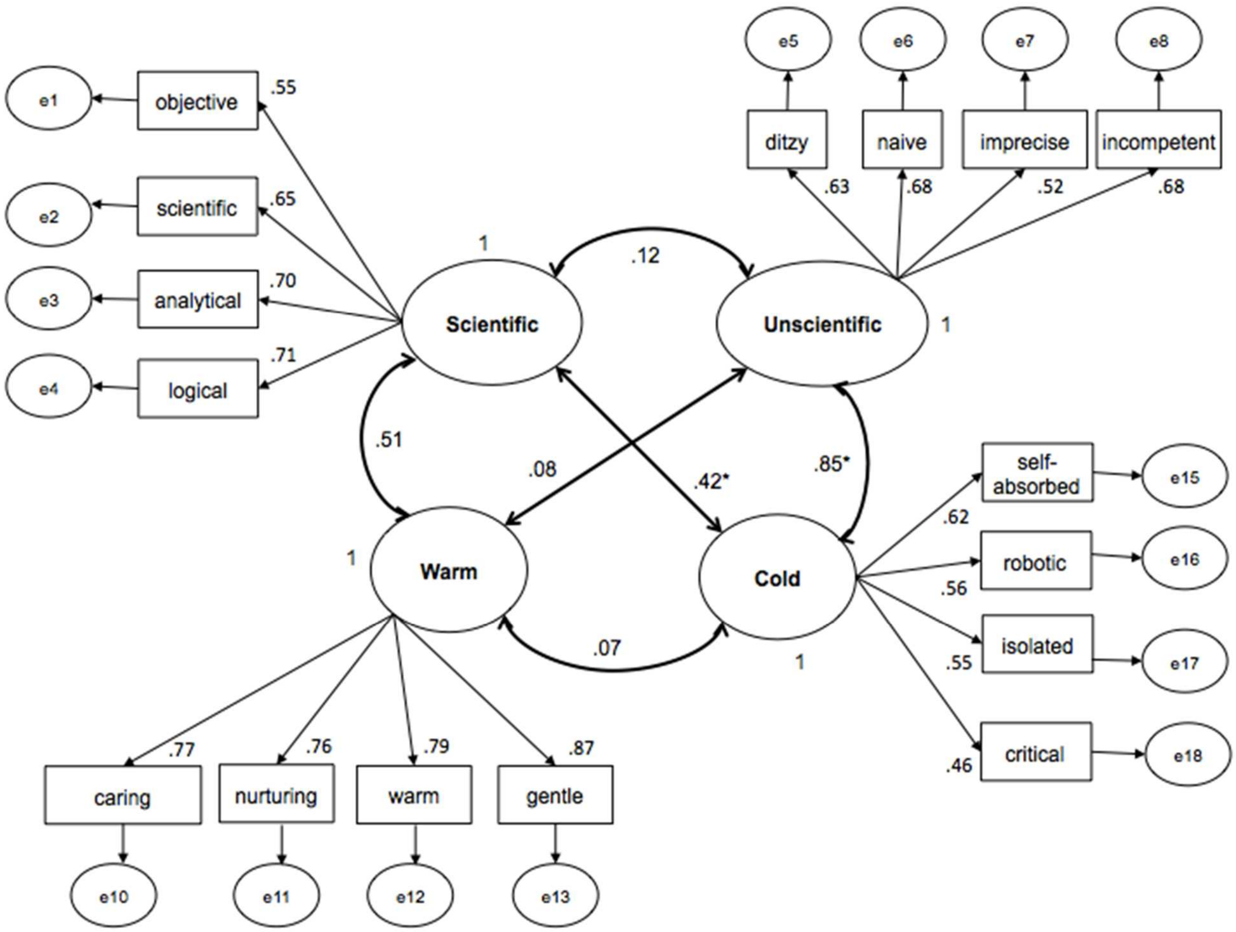


Figure 18. Standardized CFA for traits for male targets.

An analysis of gender invariance was again conducted, fixing the items with the highest loadings for each factor to 1. There was evidence for gender invariance for both female and male targets (Tables 28 and 29), meaning that men and women interpreted the traits, their underlying factors, and the relationship between factors, in a similar way.

Model	χ^2(df)	Change χ^2	*CFI*	*RMSEA*
Unconstrained	366.18(196)	--	.88	.07
Add fixed loadings	381.09(208)	14.91(12)	.87	.07
Add fixed factor variances	382.75(212)	1.66 (4)	.88	.07
Add fixed latent factor covariances	392.30(218)	9.55(6)	.87	.07

Table 28. Gender-science stereotype invariance for female targets. All *p* values are non-significant for the chi-square change.

Model	χ^2(df)	Change χ^2	*CFI*	*RMSEA*
Unconstrained	338.04(196)	--	.91	.07
Add fixed loadings	349.08(208)	11.04(12)	.91	.07
Add fixed factor variances	353.20(212)	4.12(4)	.91	.07
Add fixed latent factor covariances	365.22(218)	12.02(6)+	.91	.07

Table 29. Gender-science stereotype invariance for male targets. +Marginal change in chi-square, p = .061, for adding latent factor covariances.

In sum, the gender ideology and trait stereotype scales effectively measured the constructs of interest as intended, and they did so similarly for male and female participants. Despite a strong latent correlation between unscientific and cold traits, model comparisons confirmed that all four factors were needed to capture trait stereotypes.

Test Re-Test Reliability

I next examined how stable the explicit ratings were from pre-screen to the lab, partialing participant gender and condition. Although below conventional thresholds for good test re-test reliability (see Table 29), the correlations seemed decent given the amount of time in between the two sessions (between 1-16 weeks), as well as the variety of disparate measures included in both sessions (in the pre-screen, other measures included by other experimenters; in the laboratory session, primarily the GNAT).

Scale	Correlation
Blind	.55**
Aware	.55**
Assimilation	.65**
Segregation	.61**
Scientific Women	.46**
Scientific Men	.39**
Unscientific Women	.54**
Unscientific Men	.53**
Warm Women	.53**
Warm Men	.59**
Cold Women	.39**
Cold Men	.47**
Thermometer Women	.50**
Thermometer Men	.36**
Personal STEM stereotype	.60**
Societal STEM stereotype	.44**

Table 30. Test re-test reliabilities for explicit measures. Participant gender and condition are partialed. ** $p < .001$.

Evaluation

Explicit Evaluation.

Thermometer ratings for men and women were regressed onto participant gender. This analysis revealed a main effect of Target Gender, F(1, 345) = 23.72, p < .001, that was qualified by Participant Gender, $F(1, 345) = 111.62$, $p < .001$. Warmth depended on participant gender: female participants felt warmer towards men than male participants, $F(1, 345) = 59.93$, $p < .001$, and male participants felt warmer towards women than female participants, $F(1, 345) = 27.75$, $p < .001$. In fact, females reported greater warmth towards men than towards women (thermometer difference = 7.10, $F(1, 345) = 19.01$, $p < .001$), and males reported just the opposite (thermometer difference = 19.24, $F(1, 345) = 103.84$, $p < .001$).

The next analysis explored whether gender ideology moderated gender evaluation. Each of the gender ideologies (mean centered) and their interactions with participant gender was added as a predictor. The Target and Target × Participant Gender effect described above persisted, but there were also two three-way interactions: Target Gender × Participant Gender ×

Segregation, $F(1, 337) = 5.94$, $p = .01$, and Target Gender × Participant Gender × Assimilation, $F(1, 337) = 9.35$, $p < .01$. Looking at simple effects within participant gender, segregationism and assimilationism did not moderate warmth towards men and women for females, *p*s > .23. For males, however, both assimilationism and segregationism predicted a weaker evaluative preference for women, as evidenced by Target × Assimilation and Target × Segregation interactions, $F(1, 337) = 9.16$, p < .01, and $F(1, 337) = 5.64$, $p = .02$, respectively. At least for males, this supports the four-fold table of gender ideology—assimilationism and segregationism are related to a more negative evaluation of women.

Implicit evaluation.

Recall that participants in the GNAT completed four blocks that assessed their associations of men and women with good nouns (e.g., sunset, puppy) and bad nouns (e.g., vomit, cockroach). A 2 (Target: Men s. Women) × 2 (Valence: Good vs. Bad) × 2 (Condition: Name vs. Photos) × 2 (Participant Gender) ANOVA was conducted with repeated measures on the first three factors.[13] A Target Gender × Valence interaction indicated that good and bad words were differentially associated with men and women, $F(1, 223) = 99.59$, $p < .001$; simple analyses within valence revealed a large "women are wonderful" effect such that women were much more strongly associated with good words than were men, $F(1, 223) = 86.01$, $p < .001$, and men were much more strongly associated with bad words than women, $F(1, 223) = 39.51$, $p < .001$. This depended on participant gender, however (i.e., there was a Target Gender ×

[13] This analysis did not include the Low Prototypic condition; due to a programming error, one block within that condition (women + good words) had a response deadline of 500 milliseconds instead of 600 milliseconds, which likely artificially deflated d' within that block compared to other blocks. Because there were no substantive differences between Low Prototypic and High Prototypic stimuli with regards to stereotypes, it seemed reasonable to remove the Low Prototypic condition from the analysis with evaluation. Moreover, a 2(Target Gender) x 3(Condition) x 2(Participant Gender) ANOVA for bad associations showed no difference between Low Prototypic and High Prototypic photos, $p = .61$, nor any interactions of condition with participant gender.

A second programming error occurred in specifying the order In which blocks were presented. In one of the12 counter-balanced block presentation orders (n = 32), the men + good was presented twice, and men + bad was not presented. The second completion of "men + good" was omitted for these 32 participants and the "men + bad' trials were missing, therefore degrees of freedom are deflated for these analyses.

Valence × Participant Gender interaction), $F(1, 223) = 45.18$, $p < .001$, such that female participants had a stronger evaluative bias (see Figure 19). The Target Gender × Valence interaction remained significant within males, indicating that they had same pattern of evaluative associations with the genders as females, $F(1, 223) = 4.71$, $p = .03$. However, looking at the simple analyses within valence, males only marginally associated men with bad words more than women, $F(1, 223) = 3.49$, $p = .06$, and they did not associate women with good more than men with good, $F(1, 223) = 1.86$, $p = .17$. Thus evaluative biases preferring women were much weaker for males than females.

Whether participants were exposed to High Prototypic photos or generic names did not affect their evaluative bias, $p = .66$, nor did it interact with participant gender, $p = .11$. Thus, stimuli did not affect preference for women over men. Mirroring results with stereotypes, average d' was again faster in the High Prototypic than the Generic Names Condition, $F(1, 223) = 62.48$, $p < .001$; this did not depend on participant gender, and male and female participants responded with equal sensitivity, $p = .99$.[14]

---

[14] Performing the same analysis but only for d' to word stimuli showed equivalent sensitivity in the stereotypic and generic name condition, $F(1, 224) = 1.45$, $p = .23$; thus in contrast to the findings with implicit stereotypes, for the evaluative GNAT blocks, the greater sensitivity in the stereotypic condition relative to the generic name condition was largely driven by gender stimuli, not word stimuli.

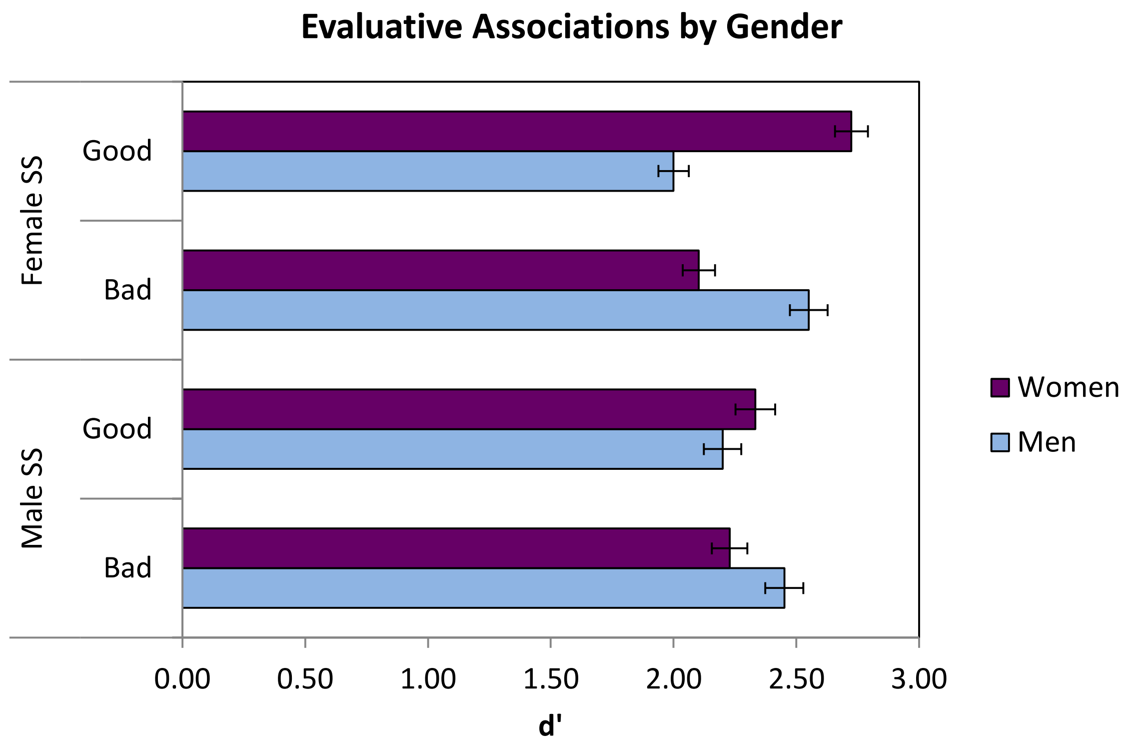


Figure 19. Implicit evaluation of men and women by participant gender.

Each gender ideology was then added to the analysis above as a mean-centered predictor, along with its interaction with participant gender. The same “women are wonderful” effect described above persisted, and again it was stronger for female participants. In addition, the “women are wonderful effect” was moderated by Assimilation, as evidenced by a Target Gender × Valence × Participant Gender × Assimilation interaction, $F(1, 203) = 4.38$, $p = .04$. Looking within participant gender, Assimilation was significant for male participants, such that the Target × Valence interaction was weaker as Assimilation increased, $F(1, 203) = 4.69$, $p = .03$. That is, males who more strongly endorsed assimilation showed less of “women are wonderful” effect. Assimilation had no effect for female participants, $p = .44$.

Next, the relationship between the evaluative GNAT (an indirect measure of evaluation) and thermometer (a direct measure of evaluation) was examined to assess the correspondence between implicit and explicit measures of evaluation. Partialing out participant gender and condition, the thermometer difference between men and women at pre-screen was unrelated to the implicit “women are wonderful effect”, *r*(206) = -.07, *p* = .34. Yet again, there was no evidence of a meaningful relationship between implicit and explicit evaluations of the genders. Finally, the High Prototypic photo condition did not differentially affect evaluation relative to generic names, although overall sensitivity was greater in the photo condition than the generic name condition.

Study 2 Discussion

Study 2 found evidence for both implicit and explicit gender-science stereotypes among males and females alike. First, a pre-test established the positive and negative traits stereotypic of scientists (relative to ECE); in accordance with the *Science Stereotypes Hypothesis*, both scientific and cold traits were judged as more descriptive of scientists than ECE, whereas both unscientific and warm traits were judged as more descriptive of ECE than scientists. Second, in accordance with the *Gender-science Stereotypes Hypothesis*, undergraduate participants estimated that more men possessed scientist stereotypic traits (i.e., scientific and cold) than women, and that more women possessed scientist counter-stereotypic traits than men (i.e., unscientific and warm).[15] These same stereotypes each emerged at an implicit level among the same participants, as revealed in a go/no-go association task pairing men and women each with the scientific, unscientific, warm and cold traits.

Participant gender differences emerged for both implicit and explicit stereotypes, although with a slightly different pattern for each. Although implicit stereotypes were present for both males and females, males and females showed an implicit in-group bias: compared to males, females possessed *stronger* implicit warm and cold stereotypes, and *weaker* implicit

[15] Significant unscientific stereotypes did not emerge in the pre-screen, but did emerge in the lab session percent estimate task. This was the case whether *unscientific* was defined using the same traits as in the pre-screen (which were slightly different from those later selected for use in the GNAT) or the identical traits used in the pre-test and GNAT. This suggests the idiosyncratic trait differences used to represent *unscientific* do not entirely explain why the pre-screen did not detect a reliable unscientific stereotype.

scientific stereotypes. This in-group bias did not emerge in explicit stereotypes, where females actually expressed *weaker* warm and cold stereotypes than males, but were equivalent to men in endorsement of scientific stereotypes. It may be that in-group biases are purposefully adjusted in explicit ratings.

The weakest stereotype index both implicitly and explicitly was unscientific. One potential reason this stereotype was weaker than the others is that unscientific traits were also rated as the least favorable traits in the pre-test (M = 2.10, SD = .99) compared to cold (M = 2.51, SD = 1.04), scientific (M = 5.59, SD = .81), and warm (M = 6.41, SD = .67). The negativity of these traits may have made participants especially reluctant to ascribe them to social groups compared to the other traits. Stronger stereotypes emerged on the warm/cold dimension than on the scientific/unscientific dimension for both implicit and explicit stereotypes, with warm stereotypes especially strong. This aligns with a body of research indicating that communal stereotypes are especially stable over time (Abele, 2003; Croft, Schmader, & Block, 2015; Diekman et al., 2010; Twenge, 1997), and implies that barriers to women's success in STEM may very well be based as much or more in their incongruity with the perceived communal aspects of STEM careers (or lack thereof), rather than the agentic aspects that align more with perceptions of scientific ability (Ceci & Williams, 2011; Ceci et al., 2009; Diekman et al., 2011).

Moderation by Gender Prototypicality

There was no support for the *Gender Prototypical Stimuli Hypothesis,* that implicit gendered science traits stereotypes would be stronger when presented with high prototypical men and women than low prototypical men and women. Implicit stereotypes were largely consistent across condition, regardless of whether the men and women presented in the GNAT were high or low on prototypicality. One unexpected Participant Gender × Condition (High Prototypic vs. Low Prototypic) interaction emerged for unscientific traits, indicating that for female participants, implicit unscientific stereotypes were in fact *stronger* in the Low Prototypic condition than the High Prototypic condition. Indeed, females in the High Prototypic condition (i.e., those who saw feminine women and masculine men in the GNAT) showed no evidence of implicit unscientific stereotypes. Speculatively, this may have been due to the Low Prototypic men being perceived as nerdy, making them less associated with unscientific traits. As seen in Figure 14, High Prototypic men were more associated with unscientific to the same extent as High Prototypic women were, but Low Prototypic men were especially weakly associated with unscientific (and less so than Low Prototypic women). That said, on balance there were largely no effects of the gender prototypicality manipulation.

Although warm stereotypes were prevalent across all conditions, they were even stronger in the generic names than the photo conditions. This aligns with previous research showing that abstract names can elicit stronger bias than pictures (Mitchell et al., 2003, Experiment 5; Nosek et al., 2002). It is ambiguous, however, why the name vs. photo effect only emerged for warm trait associations.

Condition most clearly influenced overall performance on the implicit task (i.e., ability do discriminate signal from noise). Specifically, participants showed greater sensitivity to photos than to names, likely due to greater efficiency in processing faces than words (Farah, 1992; Glaser & Glaser, 1989; Israel & Schacter, 1997; Mitchell et al., 2003). They also showed greater sensitivity to High Prototypic photos than Low Prototypic photos, likely due to the greater fluency of prototypical stimuli relative to non-prototypical stimuli (Winkielman, Halberstadt, Fazendeiro, & Catty, 2006), and the greater ease with which the High Prototypic photos were likely categorized as male or female (Macrae & Martin, 2007).

Gender Ideology

Replicating previous research on the gender ideology scale (Hahn et al., under review), the positive perspectives (gender awareness and gender blindness) were more strongly endorsed than the negative perspectives (assimilationism and segregationism), especially among female participants. Specifically, females were more likely to support gender blindness and less likely to support assimilationism and segregationism than males. Males and females supported gender awareness equivalently. Despite mean differences on the ideologies, a CFA importantly demonstrated measurement invariance across gender, meaning the items were interpreted similarly across males and females and that the ideologies related to one another in a similar way.

Research concerning ethnic ideology also suggests that majority and minority group members maintain somewhat different ideologies. In the United States, minority group members typically endorse colorblindness to a lesser extent than majority group members (Ryan, Casas, & Thompson, 2010; Ryan, Hunt, Weible, Peterson, & Casas, 2007), who may use colorblindness as a strategy to avoid appearing biased or expressing racial bias (Apfelbaum, Pauker, Sommers, & Ambady, 2010). Moreover, majority group members in the United States endorse colorblindness and multiculturalism somewhat equally (Morrison & Chung, 2011; Ryan et al., 2007), whereas in Europe, majority group members prefer colorblindness over multiculturalism and endorse multiculturalism less than minority group members (Verkuyten, 2005). This may be due to cultural differences in Europe and the United States—whereas the US features relatively little national discussion or debate about diversity ideologies, discourse concerning the best diversity approach and comparing and contrasting approaching across countries is more prevalent in Europe (Rattan & Ambady, 2013).

Whereas the present research suggests that women (the subordinate group) prefer gender blindness relative to men, research on ethnic ideology primarily indicates that ethnic minorities (also the subordinate group) prefer multiculturalism relative to ethnic majorities. This suggests not only that there is not a single clear “best” approach to diversity, but also that the prevailing ideology is malleable depending on the context and the social groups considered. For example, much more than inter-ethnic dynamics, gender dynamics entail frequent daily intergroup interactions, intergroup dependence and also complementary roles (Koenig &

Richeson, 2010). It is likely that ideologies depend on the cultural milieu associated with various countries (Guimond et al., 2013), workplaces (Plaut et al., 2009) or time periods (Plaut, 2010), which suggests that they are malleable and could be manipulated to create more welcoming environments.

Moderation by Gender Ideology

There was some support for the *Gender Ideology Hypothesis,* that perceiver gender ideologies would moderate gender-science stereotypes. Indeed, gender blindness predicted weaker scientific and unscientific stereotypes, whereas assimilationism predicted stronger scientific and unscientific stereotypes. Moreover, both segregationism and assimilationism positively predicted personal endorsement of the stereotype that men are better at math and science than women. Unexpectedly, gender blindness, despite relating to gender-science stereotypes, gender blindness was unrelated to gender-STEM stereotypes. It is not clear why these different patterns emerged. Interestingly, there was no effect of gender awareness on stereotype endorsement, and none of the gender ideologies affected warm or cold trait stereotypes.

There was minimal evidence that the ideologies affected *implicit* gender-science stereotypes. A small Participant Gender × Assimilationism interaction for the implicit unscientific index revealed that for male participants only, assimilationism was related to a stronger implicit unscientific stereotype. This suggests that perhaps the GNAT captured meaningful individual differences in gender ideologies between participants, but this finding would need to be replicated in future work.

Turning to evaluations, assimilationism and segregationism predicted weaker explicit preference for women over men for males only, and gender awareness and gender blindness had no impact on participant’s general evaluation of women and men. However, gender blindness ameliorated an evaluative preference for traditional men and women (i.e., male politicians and scientists, stay at home moms and female nurses) relative to non-traditional men and women (i.e., female politicians and scientists, stay at home dads and male nurses), whereas segregationism and assimilationism both exacerbated an evaluative preference for traditional targets compared to non-traditional targets. The single effect of ideologies on *implicit* evaluations showed that for males only, greater assimilationism was related to weaker associations of women as good, relative to men as good. Overall, this pattern of results suggests that for men in particular, assimilationism and segregationism are system-justifying ideologies that seek to avoid the social change that would bring about gender equality. Future research should investigate this hypothesis.

The present research is novel in examining the effect of chronic ideologies on stereotyping and bias, particularly implicit stereotyping and bias. Most research regarding the downstream consequences of intergroup ideologies has manipulated the ideologies rather than measuring them (e.g., Apfelbaum et al., 2010; Correll, Park &, Smith, 2008; Gutiérrez & Unzueta, 2010; Wolsko et al., 2000, 2006). In the single previous study I am aware of concerning the relationship between ideology and implicit bias, Richeson and Nussbaum (2004) exposed participants to a persuasive article arguing for the superiority of either colorblindness or multiculturalism as strategies to intergroup relations. Although a strong implicit pro-white bias was present in both conditions (i.e., more positive attitudes towards Whites than Blacks), exposure to the colorblind message exacerbated this bias on both explicit and implicit measures. Importantly, they found that the manipulations had an effect over and above personal agreement with the perspectives (the authors did not report whether personal agreement with the perspective affected implicit or explicit bias). Future research should examine the consequences of manipulating gender ideologies, although successfully inducing the two negative perspectives may be challenging given people generally do not subscribe to them.

The (Lack of) Relationship Between Implicit and Explicit Measures

Finally, the *Implicit-Explicit Correlation Hypothesis* explored the relationship between implicit and explicit attitudes. Despite a similar and clear meal-level pattern of implicit and explicit stereotypes, there was no evidence for correlations between stereotypes or evaluations. Extant theories propose two explanations for the lack of relationship between implicit and explicit attitudes: 1) implicit and explicit measures are assessing different constructs (Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000), and 2) implicit and explicit attitudes reflect the same underlying construct, but implicit attitudes assess an initial, immediate reaction that is then adjusted based on deliberate and intentional thoughts. Thus, measures that assess these two types of attitudes will diverge to the extent that people intentionally and consciously adjust their explicit attitude (Fazio & Olson, 2003). In other words, the two attitudes will differ to the extent that people have the motivation and opportunity to adjust their explicit responses.

Typically, there is a significant relationship between implicit and explicit attitudes, although its size ranges considerably (Hofmann et al., 2005; Nosek, 2005). Using a large sample from Project Implicit, Nosek (2005) found that the typical correlation was .36, although this depended on several moderators. Specifically, the correlation between implicit and explicit valuation was smaller with: 1) greater self-presentation concern, 2) weaker evaluations of the objects, 3) decreased dimensionality (the degree to which the target objects conform to a simple, bipolar structure), and 4) increased normativity of the attitude. All four of these moderators may have contributed to the lack of correlation in the present research: participants may have experienced self-presentation concerns, had relatively weak gender-science stereotypes, and perceived that their attitudes were widely held and normative. Moreover, it is not clear that the stereotypes and evaluations assessed conform to a simple, bipolar structure, which may also have contributed to a lack of a relationship.

Hofmann et al. (2005) conducted a meta-analysis of the relationship between the implicit association test and explicit self-report measures, examining a variety of potential moderators (e.g., general research domain, type of target stimuli, type of self-report measure). They found a typical correlation of .24 between self-reported measures and those assessed by the IAT, although this ranged between .001 and .471. Several of their findings shed light on the lack of correlation between the implicit and explicit stereotype measures in Study 2. First, correlations were stronger the more that self-reports on the topic tend to be spontaneous (i.e., people rely more on their gut reactions when asked to report their attitudes on the given topic, as assessed by judges). Second, the average correlations for stereotyping research were smaller than those found for consumer research and group attitudes, perhaps due to self-presentational concerns. Third, affective self-report measures (e.g., thermometer ratings) showed stronger correlations than cognitive self-report measures (e.g., trait ratings). Fourth, correlations were lower for scales (aggregate measures of several items, like those used in the present work to assess the various stereotype indices) than for semantic differential or trait ratings. Correlations were also stronger when the IAT employed evaluative noun stimuli compared to evaluative adjectives (such as those employed in the stereotyping GNAT) or thematic word stimuli. Finally, relative self-report measures (e.g., measures that directly compare two groups), showed higher correlations than absolute judgments (e.g., measures that included ratings of only one target group). However, computed difference scores, like those used in the present research for stereotype indices, also showed higher correlations than absolute judgments.

This work suggests that lack of correlation between implicit and explicit stereotypes may have been due to being a stereotyping task, employing explicit measures that were likely non-spontaneous (i.e., the percent estimate task), self-presentational concerns on self-report measures, and the use of evaluative adjectives rather than evaluative nouns. However, Study 2 also found no relationship between implicit and explicit evaluations (e.g., thermometer ratings);

it is not clear why this is the case given that this task included many facets that typically increase the correlation between implicit and explicit attitudes.

Ancillary Analyses

Ancillary analyses indicated that the explicit measures possessed good measurement properties. CFAs supported that the gender ideologies and the gendered science traits showed decent fit and measurement invariance across men and women, meaning that males and females interpreted the items in the same way and the factors related to one another similarly. Moreover, many theoretically meaningful relationships emerged between the explicit measures. For example, explicit scientific and unscientific stereotypes corresponded to stronger personal STEM stereotypes, stronger hostile sexism, and weaker internal motivation to respond without sexism. The two positive ideologies were both positively related to internal motivation to respond without sexism. Gender blindness was negatively related to hostile sexism as well as greater warmth for traditional vs. non-traditional gender roles, whereas gender awareness was positively related to benevolent sexism. Assimilationism and segregationism were related to other variables in the same way: both predicted stronger hostile and benevolent sexism, weaker internal motivation and stronger external motivation, and finally, greater warmth towards people in traditional gender roles than non-traditional gender roles. Notably, there was no effect of gender awareness on stereotype endorsement. Indeed, gender awareness was the ideology with the least predictive value. It related positively only with benevolent sexism as well as internal motivation to respond without sexism. Thus gender awareness seems to be situated in an interesting position of genuinely striving to be non-sexist, while also celebrating women as different from men.

On the other hand, implicit stereotypes were related to any explicit measures, whether assessed in the pre-screen or in the lab session. This was somewhat surprising given previous showing that, for example, people low on external and high on internal motivation to respond without prejudice showed weaker implicit racial bias (Devine, Ashby, Amodio, Harmon-Jones, & Vance, 2002). A parallel exploratory analysis conducted in the present research found no evidence for such a relationship (i.e., between motivation to respond without sexism and implicit stereotypes or evaluation).

Ancillary analyses also examined implicit and explicit evaluation. Aligning with previous research showing people have more positive attitudes towards women than men (Eagly & Mladinic, 1989; Eagly, Mladinic, & Otto, 1991), the “women are wonderful” effect emerged on average across males and females on both implicit and explicit (thermometer) evaluations. Importantly however, this overall preference for women was moderated by participant gender, and showed different patterns for the implicit vs. explicit evaluations: for explicit judgments, males and females both showed an outgroup preference; females rated men more warmly than women and males did just the opposite. For implicit ratings, on the other hand, females showed a pro-female bias. Males also associated men with bad more than women with bad, but to a lesser extent than females. Finally, males did not associate women with good more than men with good. This may reflect that men and women may have purposefully adjusted their explicit evaluations of the opposite sex so as to appear that they did not have an in-group preference.

In sum, men and women alike possessed explicit and implicit gender-science stereotypes. Implicit stereotypes were generally quite robust regardless of gender instantiation and gender ideology, whereas explicit stereotypes—particularly for scientific and unscientific judgments—were more susceptible to one’s gender ideology and other explicit gender-related beliefs.

CHAPTER 6. General Discussion

The primary goal of this research program was to examine how gender ideologies affect gender stereotyping and bias. In seeking to address this question, several other interesting findings emerged: more feminine female (but not male) scientists were judged as less likely to be scientists (Study 1), people associated scientific traits with men more than with women both implicitly and explicitly (Study 2), and implicit stereotypes and evaluations were persistent and strong regardless of target gender prototypicality, perceiver gender ideology, or other sociopolitical beliefs that could theoretically influence implicit attitudes (e.g., ambivalent sexism; Study 2). In contrast, explicit gender-science stereotypes—in particular scientific and unscientific stereotypes—were indeed moderated by gender ideologies.

Contributions

One contribution of the current research is extending research regarding gender stereotypes and STEM to the actual traits ascribed to men and women on both an implicit and explicit level. Existing literature has shown that people implicitly associate math with men more than women; whereas these associations may simply reflect one's exposure to women in STEM domains (Miller, Eagly, & Linn, 2014), willfully assigning fewer scientific stereotypes to women than to men may reflect a more problematic phenomenon wherein women are actually viewed as lacking the qualities required to succeed in science. It is one thing to think *science* and think *male*, and another to think *scientific* and think *male*. Gender-science stereotypes also emerged at an explicit level, exhibiting even larger effect sizes than implicit gender-science stereotypes for all dimensions except *unscientific* (see Table 21). This may because people feel less guilt expressing sexism relative to other biases, such as racism, and because people widely believe that men and women do and *should* possess characteristics that complement one another (Czopp & Monteith, 2003; Glick & Fiske, 1996; Prentice & Carranza, 2002).

According to role incongruity theory, the perceived incongruity between the traits that women possess and those demanded by the career scientist will likely result in bias and discrimination (Eagly & Karau, 2002). For example, women may be judged less favorably than men for scientist positions (e.g., Moss-Racusin et al., 2012), may struggle to access and succeed in scientist positions, and when performing scientist behaviors, may be evaluated less favorably than men. Moreover, contextual circumstances that enhance the perceptions of incongruity between the female gender role and the scientist role, such as feminine physical appearance or becoming a mother, should exacerbate the perceived mismatch between the female gender role and the role scientist. Indeed, Study 1 indicated that female scientists with more feminine facial appearance were deemed (by naïve participants) less likely to be scientists than their masculine counterparts. Future research is necessary to understand the precise mechanism underlying this finding.

Limitations & Future Directions

Although femininity gender prototypicality influenced judged likelihood of being a scientist, it had a negligible impact on implicit stereotypes (Study 2). In retrospect, the GNAT was likely not the optimal method to examine the impact of stimulus variance. The GNAT, like the IAT, requires categorizing the stimuli as either “men” or “women,” shifting attention away from each individual stimulus. Indeed, there is some evidence that responses on an IAT are driven more by attitudes towards the category labels than the stimulus items themselves (De Houwer, 2001). Moreover, when stimuli affect responses on the IAT, it is still allegedly via categorization. For example, Govan and Williams (2004) showed that they could reverse a pro-White IAT bias by instantiating the category White with negative exemplars such as Adolph Hitler and Charles Manson, and instantiating the category Black with positive exemplars such as Eddie Murphy and Michael Jordan. They posited that the stimuli lead participants to implicitly redefine the category labels (e.g., “bad White people”), resulting in diminished or reversed implicit biases. In the present research, it seems unlikely that participants subtyped the stimuli as "masculine women" or "feminine women" for two reasons: variations in femininity were quite subtle, and participants were not provided with the entire range of stimuli that would make some faces particularly feminine and others particularly masculine.

Future research should employ an implicit task in which participants are not required to categorize the stimuli, such as a Lexical Decision Task (Wittenbrink et al., 1997) or a Sequential Priming Task (Fazio, Jackson, Dunton, & Williams, 1995; see also Ito et al., 2011). Such tasks would offer two other important benefits. First, gender prototypicality could be manipulated *within* participants rather than between them; in the Sequential Priming task, each participant could be exposed to a range of faces along the feminine/masculine spectrum, rather than simply highly feminine or masculine faces as was done in Study 2. In the present research, it may have been problematic to divide the stimuli between participants such that they were only exposed to a limited range of gender prototypical individuals; Lick and Johnson (2014) demonstrated that perceptual gender norms are quickly and automatically calibrated based on recent exposure to gendered facial features. This suggests that participants in Study 2 may have quickly adapted to the faces in the GNAT and/or may not have interpreted the faces as highly feminine or masculine in the absence of relative context from other faces (Kenrick & Gutierres, 1980; Lick & Johnson, 2014). Second, such a design would allow one to examine prototypicality as a continuous predictor rather than a categorical one, increasing the power of the design (Irwin & McClelland, 2003).

Moreover, there is evidence that category-based implicit measures, or implicit tasks in which participants have to categorize each stimulus (the IAT) actually tap a different construct from those in which participants process each individual stimulus or exemplar (the Sequential Priming Task). For example, Olson and Fazio (2003) found that implicit attitudes as assessed by the IAT and Sequential Priming Task were not only uncorrelated with each other, but showed different magnitudes of bias (with the IAT showing stronger bias than the Sequential Priming Task), suggesting that they are measuring distinct underlying constructs (Olson & Fazio, 2003).

In addition to employing a different implicit measure, another option would be to measure explicit judgments while still assessing the extent to which stereotyping based on feminine features occurs via an automatic process. For example, Blair and colleagues showed that targets with more Afrocentric facial features exacerbated Black stereotypes in an automatic manner (Blair et al., 2004b). In their study, participants read four short biographies, each of which embodied either positive or negative stereotypic or counter-stereotypic information pertaining to Blacks or Whites. For each profile, participants were exposed to a variety of Black and White faces and asked to estimate the likelihood that the individual was the person described in the biography. Results showed that increasingly Afrocentric individuals were judged as more likely to be the positive and negative Black stereotypic individual. In subsequent versions of the study, constraints were introduced to assess the extent to which Afrocentricity influenced judgments automatically. First, Afrocentricity still influenced stereotypes when participants' cognitive resources were restricted, indicating that the bias was operating efficiently. Moreover, in contrast to stereotypes based on racial categories, participants were relatively unaware of and unable to control their Afrocentricity bias, even after being clearly informed about the bias. Adopting a similar paradigm to test how feminine appearance influences gender-science stereotypes—and whether it does so differently for people who subscribe to different gender ideologies—would be useful in future research. Indeed, Ko and colleagues used a similar paradigm to show that increasing vocal femininity exacerbated both positive and negative gender stereotypes (Ko, Judd, & Blair, 2006)

Implications

The present work suggests that implicit stereotypes may not be responsive to one's explicit beliefs and values, including one's explicit endorsement of stereotypes, perspectives on how to approach gender differences, ambivalent sexism, or internal and external motivation to respond without sexism. It may well be that implicit associations are more affected by one's accumulated experiences and exposure to women in STEM. Extant research shows that exposure to a greater number of women in STEM can reduce implicit stereotypes of men with math more than women with math (Nosek et al., 2009; Miller et al., 2015); it would be worth examining whether implicit-gender STEM stereotypes also depend on one's exposure to women in STEM.

Given that explicit gender-science stereotypes showed equivalent, if not stronger, effect sizes in comparison to implicit gender-science stereotypes, and that they depended on one's gender ideologies and other explicit gender-related beliefs, shifting gender ideologies and/or explicit gender-science stereotypes may initially be more plausible than shifting implicit stereotypes. Indeed, scientific and unscientific stereotypes depended on gender blindness and assimilationism, suggesting that one potential intervention in STEM fields would be to manipulate these gender ideologies and/or to showcase that women do not need to be exactly like men in order to be successful (i.e., they do not need to assimilate). This seems particularly important given research showing that the ideologies endorsed by the dominant group in a workplace have a tangible impact on minority engagement and perceived bias. Plaut et al.'s (2009) compelling study showed that Whites' endorsement of *assimilation* (e.g., "'Employees should downplay their racial and ethnic differences'' and ''The organization should encourage racial and ethnic minorities to adapt to mainstream ways'') increased their minority co-workers' sense of racial bias and depleted their workplace engagement, whereas Whites' endorsement of multiculturalism had the opposite effect (note that Plaut et al., labeled this perspective colorblindness, but it has a distinctly assimilationist message). Future research should explore how prevailing workplace gender ideologies affect women's perceived sexism and engagement.

In male-dominated fields, promoting a gender blind ideology that advocates acknowledging that each individual is different regardless of his or her gender may be a promising strategy to foster a more welcoming environment. Even a gender aware message that highlights that men and women might approach tasks differently, but in an equally valuable way, might beckon more women into STEM. Indeed, women's increasing presence in leadership roles has elicited discussion of how women may offer a more socially engaged, egalitarian leadership style relative to men (Eagly & Carli, 2003). Together, these sets of studies suggest that both implicit and explicit gender-science stereotypes exist, and that the latter are related to one's gender ideologies and other gender-related explicit beliefs. Future research is certainly needed to determine whether the ideologies in fact *cause* changes in gender-science stereotypes. Moreover, at least in the present research, gender prototypical appearance largely did not influence implicit gender-science stereotypes, although more feminine women were explicitly deemed less likely to be scientists. The interface between gender ideology and tendencies or willingness to use facial femininity as a cue to scientific ability should also be examined.

Overall, this research suggests that gender ideology offers an additional tool that might further women's attraction to, retention, and well-being in STEM. Given that scientists who embody STEM stereotypes are discouraging to men and women alike (Cheryan, Plaut, Handron, & Hudson, 2013; Cheryan et al., 2011; Hannover & Kessels, 2004), certain messages may be appealing to both genders: those emphasizing that people should be appreciated for who they are as a unique individual, regardless of their gender (gender blindness), that men and women might bring different experiences to the table that make science stronger and more nuanced (gender awareness); and that being a successful scientist does not require conforming to the stereotypic image of a scientist (assimilationism) or else abandoning the field (segregationism). In sum, gender ideology may offer an underutilized lens through which to better understand and address the scarcity of women in STEM. As has been seen in the ethnic and racial domain, ideologies about how to approach social group differences are important predictors of stereotyping, bias, belonging and engagement. It is past time that we harness the power of intergroup ideologies to ameliorate stereotypes and improve women's outcomes in male-dominated fields such as STEM.

References

Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology*, *85*(4), 768–776. http://doi.org/10.1037/0022-3514.85.4.768

Apfelbaum, E. P., Pauker, K., Sommers, S. R., & Ambady, N. (2010). In Blind Pursuit of Racial Equality? *Psychological Science*, *21*(11), 1587 –1592. http://doi.org/10.1177/0956797610384741

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. http://doi.org/10.1016/j.jml.2007.12.005

Bafumi, J., & Gelman, A. E. (2006). Fitting Multilevel Models When Predictors and Group Effects Correlate. Retrieved from http://academiccommons.columbia.edu/item/ac:125243

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, *68*(2), 181–198. http://doi.org/10.1037/0022-3514.68.2.181

Banchefsky, S. & Park, B. Gender ideologies shift as a function of the proportion of men in one's academic major. In preparation.

Banchefsky, S., Westfall, J., Park., B., & Judd, C. Women with feminine facial appearance are judged as less likely to be scientists. Under review.

Bargh, J. A. (1994). The Four Horsemen of Automaticity: Intention, Awareness, Efficiency, and Control as Separate Issues. In R. Wyer & T. Srull (Eds.), *Handbook of Social Cognition* (pp. 1–1). Lawrence Erlbaum.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479. http://doi.org/10.1037/0003-066X.54.7.462

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. http://doi.org/10.1016/j.jml.2012.11.001

Bell, A., & Jones, K. (2014). Explaining fixed effects: Random effects modelling of time-series cross-sectional and panel data. *Political Science Research and Methods*.

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. http://doi.org/10.1177/1948550612440735

Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 359–374). Mahwah, NJ: Erlbaum.

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. http://doi.org/10.1002/ejsp.232

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004a). The Influence of Afrocentric Facial Features in Criminal Sentencing. *Psychological Science*, *15*(10), 674–679. http://doi.org/10.1111/j.0956-7976.2004.00739.x

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. http://doi.org/10.1037/0022-3514.87.6.763

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. http://doi.org/10.1037/0022-3514.83.1.5

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. http://doi.org/10.1177/1745691610393980

Carnevale, A. P., Strohl, J., & Melton, M. (2013). What's it worth?: The economic value of college majors. *Georgetown University Institutional Report*.

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. http://doi.org/10.1073/pnas.1014871108

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*(2), 218–261. http://doi.org/10.1037/a0014412

Cejka, M. A., & Eagly, A. H. (1999). Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment. *Personality and Social Psychology Bulletin*, *25*(4), 413–423. http://doi.org/10.1177/0146167299025004002

Chappell, B. (2015, June 10). Nobel Laureate in Hot Water for 'Trouble with Girls' in Labs. *National Public Radio.* Retrieved from http://www.npr.org/sections/thetwo-way/2015/06/10/413429407/nobel-laureate-in-hot-water-for-trouble-with-girls-in-labs

Cheng, C. (2008). Marginalized Masculinities and Hegemonic Masculinity: An Introduction. *The Journal of Men's Studies*, *7*(3), 295–315. http://doi.org/10.3149/jms.0703.295

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. http://doi.org/10.1037/a0016239

Cheryan, S., Plaut, V. C., Handron, C., & Hudson, L. (2013). The Stereotypical Computer Scientist: Gendered Media Representations as a Barrier to Inclusion for Women. *Sex Roles*, *69*(1-2), 58–71. http://doi.org/10.1007/s11199-013-0296-x

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. http://doi.org/10.1177/1948550611405218

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. http://doi.org/10.1207/S15328007SEM0902_5

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. http://doi.org/10.1016/S0022-5371(73)80014-3

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*(1), 107–112. http://doi.org/10.1177/001316447303300111

Cook, A., & Minnotte, K. L. (2008). Occupational and Industry Sex Segregation and the Work–Family Interface. *Sex Roles*, *59*(11-12), 800–813. http://doi.org/10.1007/s11199-008-9484-5

Correll, J., Park, B., & Smith, J. A. (2008). Colorblind and Multicultural Prejudice Reduction Strategies in High-Conflict Situations. *Group Processes & Intergroup Relations*, *11*(4), 471–491. http://doi.org/10.1177/1368430208095401

Crawford, M., & Unger, R. (2004). *Women and gender: A feminist psychology (4th ed.)* (Vol. xxii). New York, NY, US: McGraw-Hill.

Crisp, R. J., & Turner, R. N. (2011). Cognitive adaptation to the experience of social and cultural diversity. *Psychological Bulletin*, *137*(2), 242–266. http://doi.org/10.1037/a0021840

Croft, A., Schmader, T., & Block, K. (2015). An Underexamined Inequality Cultural and Psychological Barriers to Men's Engagement With Communal Roles. *Personality and Social Psychology Review*, 1088868314564789. http://doi.org/10.1177/1088868314564789

Czopp, A. M., & Monteith, M. J. (2003). Confronting Prejudice (Literally): Reactions to Confrontations of Racial and Gender Bias. *Personality and Social Psychology Bulletin*, *29*(4), 532–544. http://doi.org/10.1177/0146167202250923

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, *46*(5), 991–1004. http://doi.org/10.1037/0022-3514.46.5.991

De Houwer, J. (2001). A Structural and Process Analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*(6), 443–451. http://doi.org/10.1006/jesp.2000.1464

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. http://doi.org/10.1037/0022-3514.56.1.5

Devine, P. G., Ashby, E., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*(5), 835–848. http://doi.org/10.1037/0022-3514.82.5.835

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), 1051–7. http://doi.org/10.1177/0956797610377342

Diekman, A. B., Clark, E. K., Johnston, A. M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influences attraction to stem careers: Evidence for a goal congruity perspective. *Journal of Personality and Social Psychology*, *101*(5), 902–918. http://doi.org/10.1037/a0025199

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68. http://doi.org/10.1037/0022-3514.82.1.62

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the Nature of Prejudice: Automatic and Controlled Processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540. http://doi.org/10.1006/jesp.1997.1331

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist*, *50*(3), 145–158. http://doi.org/10.1037/0003-066X.50.3.145

Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *The Leadership Quarterly*, *14*(6), 807–834. http://doi.org/10.1016/j.leaqua.2003.09.004

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. http://doi.org/10.1037//0033-295X.109.3.573

Eagly, A. H., & Mladinic, A. (1989). Gender Stereotypes and Attitudes Toward Women and Men. *Personality and Social Psychology Bulletin*, *15*(4), 543–558. http://doi.org/10.1177/0146167289154008

Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are Women Evaluated More Favorably Than Men?: An Analysis of Attitudes, Beliefs, and Emotions. *Psychology of Women Quarterly*, *15*(2), 203–216. http://doi.org/10.1111/j.1471-6402.1991.tb00792.x

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*(4), 735–754. http://doi.org/10.1037/0022-3514.46.4.735

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking Deathworthy Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes. *Psychological Science*, *17*(5), 383–386. http://doi.org/10.1111/j.1467-9280.2006.01716.x

Evans, C. D., & Diekman, A. B. (2009). On Motivated Role Selection: Gender Beliefs, Distant Goals, and Career Interest. *Psychology of Women Quarterly*, *33*(2), 235–249. http://doi.org/10.1111/j.1471-6402.2009.01493.x

Farah, M. J. (1992). Is an Object an Object an Object? Cognitive and Neuropsychological Investigations of Domain Specificity in Visual Object Recognition. *Current Directions in Psychological Science*, *1*(5), 164–169.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027. http://doi.org/10.1037/0022-3514.69.6.1013

Fazio, R. H., & Olson, M. A. (2003). Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology*, *54*(1), 297–327. http://doi.org/10.1146/annurev.psych.54.101601.145225

Fine, C. (2010). *Delusions of Gender: How Our Minds, Society, and Neurosexism Create Difference*. W. W. Norton & Company.

Fine, C. (2011). Explaining, or Sustaining, the Status Quo? The Potentially Self-Fulfilling Effects of “Hardwired” Accounts of Sex Differences. *Neuroethics*, *5*(3), 285–294. http://doi.org/10.1007/s12152-011-9118-4

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychol- ogy* (4th ed., Vol. 2, pp. 357–411). New York: McGraw-Hill.

Fiske, S. T., C, J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. http://doi.org/10.1037/0022-3514.82.6.878

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. http://doi.org/10.1177/0146167292184006

Frome, P. M., Alfeld, C. J., Eccles, J. S., & Barber, B. L. (2006). Why don't they want a male-dominated job? An investigation of young women who changed their occupational aspirations. *Educational Research and Evaluation*, *12*(4), 359–372. http://doi.org/10.1080/13803610600765786

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the Processes Underlying Evaluation: Attitudes from the Perspective of the Ape Model. *Social Cognition*, *25*(5), 687–717. http://doi.org/10.1521/soco.2007.25.5.687

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*(1), 13–42. http://doi.org/10.1037/0096-3445.118.1.13

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491–512. http://doi.org/10.1037/0022-3514.70.3.491

Glick, P., Zion, C., & Nelson, C. (1988). What Mediates Sex Discrimination in Hiring Decisions ?, *55*(2), 178–186.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700–17. http://doi.org/10.1037/a0026659

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, *40*(3), 357–365. http://doi.org/10.1016/j.jesp.2003.07.002

Gray, J. (1992). Men are from Mars, women are from Venus: A practical guide for improving communication and getting what you want in your relationships. New York: HarperCollins.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley. Reprinted 1974 by Krieger, Huntington, NY.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4.

Guimond, S., Crisp, R. J., De Oliveira, P., Kamiejski, R., Kteily, N., Kuepper, B., … Zick, A. (2013). Diversity policy, social dominance, and intergroup relations: Predicting prejudice in changing social and political contexts. *Journal of Personality and Social Psychology*, *104*(6), 941–958. http://doi.org/10.1037/a0032069

Gutiérrez, A. S., & Unzueta, M. M. (2010). The effect of interethnic ideologies on the likability of stereotypic vs. counterstereotypic minority targets. *Journal of Experimental Social Psychology*, *46*(5), 775–784. http://doi.org/10.1016/j.jesp.2010.03.010

Hahn, A., Banchefsky, S., Park, B., & Judd, C. Measuring Inter-Ethnic and Inter-Gender Ideologies: Positive and Negative Aspects of Emphasizing versus Looking Beyond Group Differences. Under review.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*(3), 1369–1392. http://doi.org/10.1037/a0035028

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological*

Science in the Public Interest, *8*(1), 1–51. http://doi.org/10.1111/j.1529-1006.2007.00032.x

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*(1), 51–67. http://doi.org/10.1016/j.learninstruc.2003.10.002

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few?: women in science, technology, engineering, and mathematics*. Washington, D.C: AAUW.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. http://doi.org/10.1177/0146167205275613

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender Comparisons of Mathematics Attitudes and Affect. *Psychology of Women Quarterly*, *14*(3), 299–324. http://doi.org/10.1111/j.1471-6402.1990.tb00022.x

Irwin, J. R., & McClelland, G. H. (2003). Negative Consequences of Dichotomizing Continuous Predictor Variables. *Journal of Marketing Research*, *40*(3), 366–371. http://doi.org/10.1509/jmkr.40.3.366.19237

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*(4), 577–581. http://doi.org/10.3758/BF03214352

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.

Ito, T. A., Willadsen-Jensen, E. C., Kaye, J. T., & Park, B. (2011). Contextual variation in automatic evaluative bias to racially-ambiguous faces. *Journal of Experimental Social Psychology*, *47*(4), 818–823. http://doi.org/10.1016/j.jesp.2011.02.016

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. http://doi.org/10.1037/a0028347

Kahn, K. B., & Davies, P. G. (2011). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes & Intergroup Relations*, *14*(4), 569–580. http://doi.org/10.1177/1368430210374609

Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: Ironic effects of organizational diversity structures. *Journal of Personality and Social Psychology*, *104*(3), 504–519. http://doi.org/10.1037/a0030838

Kendall, L. (1999). Nerd nation Images of nerds in US popular culture. *International Journal of Cultural Studies*, *2*(2), 260–283. http://doi.org/10.1177/136787799900200206

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

Kenrick, D. T., & Gutierres, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, *38*(1), 131–140. http://doi.org/10.1037/0022-3514.38.1.131

Kessels, U. (2005). Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education*, *20*(3), 309–323. http://doi.org/10.1007/BF03173559

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit Stereotypes, Gender Identification, and Math-Related Outcomes A Prospective Study of Female College Students. *Psychological Science*, *18*(1), 13–18. http://doi.org/10.1111/j.1467-9280.2007.01841.x

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*. http://doi.org/10.3389/fpsyg.2013.00513

Kimball, M. M. (1994). The worlds we live in: Gender similarities and differences. *Canadian Psychology/Psychologie Canadienne*, *35*(4), 388–404. http://doi.org/10.1037/0708-5591.35.4.388

Klonis, S. C., Plant, E. A., & Devine, P. G. (2005). Internal and External Motivation to Respond Without Sexism. *Personality and Social Psychology Bulletin*, *31*(9), 1237–1249. http://doi.org/10.1177/0146167205275304

Knowles, E. D., Lowery, B. S., Hogan, C. M., & Chow, R. M. (2009). On the malleability of ideology: Motivated construals of color blindness. *Journal of Personality and Social Psychology*, *96*(4), 857–869. http://doi.org/10.1037/a0013595

Koenig, A. M., Eagly, A. H., Mitchell, A. a, & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, *137*(4), 616–42. http://doi.org/10.1037/a0023557

Koenig, A. M., & Richeson, J. A. (2010). The Contextual Endorsement of Sexblind Versus Sexaware Ideologies. *Social Psychology*, *41*(3), 186–191. http://doi.org/10.1027/1864-9335/a000026

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the Voice Reveals: Within- and Between-Category Stereotyping on the Basis of Voice. *Personality and Social Psychology Bulletin*, *32*(6), 806–819. http://doi.org/10.1177/0146167206286627

Latu, I. M., Stewart, T. L., Myers, a. C., Lisco, C. G., Estes, S. B., & Donahue, D. K. (2011). What We "Say" and What We "Think" About Female Managers: Explicit Versus Implicit Associations of Women With Success. *Psychology of Women Quarterly*, *35*(2), 252–266. http://doi.org/10.1177/0361684310383811

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. http://doi.org/10.1126/science.1261375

Lick, D. J., & Johnson, K. L. (2014). Recalibrating gender perception: Face aftereffects and the perceptual underpinnings of gender-related biases. *Journal of Experimental Psychology: General*, *143*(3), 1259–1276. http://doi.org/10.1037/a0034516

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*(1), 5–18. http://doi.org/10.1037/0022-3514.82.1.5

Losh, S. C. (2010). Stereotypes about scientists over time among US adults: 1983 and 2001. *Public Understanding of Science*, *19*(3), 372–382. http://doi.org/10.1177/0963662508098576

Macrae, C. N., & Martin, D. (2007). A boy primed Sue: feature-based processing and person construal. *European Journal of Social Psychology*, *37*(5), 793–805. http://doi.org/10.1002/ejsp.406

Macrae, C. N., Mitchell, J. P., & Pendry, L. F. (2002). What's in a Forename? Cue Familiarity and Stereotypical Thinking. *Journal of Experimental Social Psychology*, *38*(2), 186–193. http://doi.org/10.1006/jesp.2001.1496

Maddox, K. B. (2004). Perspectives on Racial Phenotypicality Bias. *Personality and Social Psychology Review*, *8*(4), 383–401. http://doi.org/10.1207/s15327957pspr0804_4

Margolis, J., Fisher, A., & Miller, F. (2000). The Anatomy of Interest: Women in Undergraduate Computer Science. *Women's Studies Quarterly*, *28*(1/2), 104–127.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442. http://doi.org/10.1006/jesp.2000.1470

Miller, D. I., Eagly, A. H., & Linn, M. C. (2014). Women's Representation in Science Predicts National Gender-Science Stereotypes: Evidence From 66 Nations. *Journal of Educational Psychology*, No Pagination Specified. http://doi.org/10.1037/edu0000005

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*(3), 455–469. http://doi.org/10.1037/0096-3445.132.3.455

Morrison, K. R., & Chung, A. H. (2011). “White” or “European American”? Self-identifying labels influence majority group members’ interethnic attitudes. *Journal of Experimental Social Psychology*, *47*(1), 165–170. http://doi.org/16/j.jesp.2010.07.019

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. http://doi.org/10.1111/j.1467-9280.2007.01995.x

Nosek, B. A. (2005). Moderators of the Relationship Between Implicit and Explicit Evaluation. *Journal of Experimental Psychology: General*, *134*(4), 565–584. http://doi.org/10.1037/0096-3445.134.4.565

Nosek, B. A. (2007). Implicit–Explicit Relations. *Current Directions in Psychological Science*, *16*(2), 65–69. http://doi.org/10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition*, *19*(6), 625–666. http://doi.org/10.1521/soco.19.6.625.20886

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*(1), 44–59. http://doi.org/10.1037/0022-3514.83.1.44

Nosek, B. A., & Smyth, F. L. (2011). Implicit Social Cognitions Predict Sex Differences in Math Engagement and Achievement. *American Educational Research Journal*, *48*(5), 1125–1156. http://doi.org/10.3102/0002831211410683

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*(26), 10593–10597. http://doi.org/10.1073/pnas.0809921106

Olson, M. A., & Fazio, R. H. (2003). Relations Between Implicit Measures of Prejudice What Are We Measuring? *Psychological Science*, *14*(6), 636–639. http://doi.org/10.1046/j.0956-7976.2003.psci_1477.x

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology*, *59*(2), 173–191. http://doi.org/10.1037/0022-3514.59.2.173

Park, B., & Judd, C. M. (2005). Rethinking the Link Between Categorization and Prejudice Within the Social Cognition Perspective. *Personality and Social Psychology Review*, *9*(2), 108 –130. http://doi.org/10.1207/s15327957pspr0902_2

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology*, *42*(6), 1051–1068. http://doi.org/10.1037/0022-3514.42.6.1051

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. http://doi.org/10.1038/29772

Plaut, V. C. (2010). Diversity Science: Why and How Difference Makes a Difference. *Psychological Inquiry*, *21*(2), 77–99. http://doi.org/10.1080/10478401003676501

Plaut, V. C., Thomas, K. M., & Goren, M. J. (2009). Is Multiculturalism or Color Blindness Better for Minorities? *Psychological Science*, *20*(4), 444 –446. http://doi.org/10.1111/j.1467-9280.2009.02318.x

Prentice, D. A., & Carranza, E. (2002). What Women and Men Should Be, Shouldn't Be, Are Allowed to Be, and Don't Have to Be: The Contents of Prescriptive Gender Stereotypes. *Psychology of Women Quarterly*, *26*(4), 269–281. http://doi.org/10.1111/1471-6402.t01-1-00066

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. http://doi.org/10.1016/S0022-1031(03)00088-X

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. http://doi.org/10.1037/0022-3514.94.4.615

Rahm, J., & Charbonneau, P. (1997). Probing stereotypes through students' drawings of scientists. *American Journal of Physics*, *65*(8), 774–778. http://doi.org/10.1119/1.18647

Rattan, A., & Ambady, N. (2013). Diversity ideologies and intergroup relations: An examination of colorblindness and multiculturalism. *European Journal of Social Psychology*, *43*(1), 12–21. http://doi.org/10.1002/ejsp.1892

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. http://doi.org/10.1162/edfp.2009.4.4.468

Rhoton, L. A. (2011). Distancing as a Gendered Barrier Understanding Women Scientists' Gender Practices. *Gender & Society*, *25*(6), 696–716. http://doi.org/10.1177/0891243211422717

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*(3), 417–423. http://doi.org/16/j.jesp.2003.09.002

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. http://doi.org/10.1037/0033-2909.110.3.574

Ryan, C. S., Casas, J. F., & Thompson, B. K. (2010). Interethnic Ideology, Intergroup Perceptions, and Cultural Orientation. *Journal of Social Issues*, *66*(1), 29–44. http://doi.org/10.1111/j.1540-4560.2009.01631.x

Ryan, C. S., Hunt, J. S., Weible, J. A., Peterson, C. R., & Casas, J. F. (2007). Multicultural and Colorblind Ideology, Stereotypes, and Ethnocentrism among Black and White Americans. *Group Processes & Intergroup Relations*, *10*(4), 617–637. http://doi.org/10.1177/1368430207084105

Schmader, T., Johns, M., & Barquissau, M. (2004). The Costs of Accepting Gender Differences: The Role of Stereotype Endorsement in Women's Experience in the Math Domain. *Sex Roles*, *50*(11-12), 835–850. http://doi.org/10.1023/B:SERS.0000029101.74557.a0

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. Available at http://www.nsf.gov/statistics/wmpd/.

Schott, G., & Selwyn, N. (2000). Examining the "Male, Antisocial" Stereotype of High Computer Users. *Journal of Educational Computing Research*, *23*(3), 291–303. http://doi.org/10.2190/V98R-5ETX-W9LY-WD3J

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. http://doi.org/10.1177/0146167203258831

Smiler, A. P. (2006). Living the Image: A Quantitative Approach to Delineating Masculinities. *Sex Roles*, *55*(9-10), 621–632. http://doi.org/10.1007/s11199-006-9118-8

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *The American Psychologist*, *60*(9), 950–8. http://doi.org/10.1037/0003-066X.60.9.950

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. http://doi.org/10.1006/jesp.1998.1373

Tenenbaum, H. R., & Leaper, C. (2003). Parent-child conversations about science: The socialization of gender inequities? *Developmental Psychology*, *39*(1), 34–47. http://doi.org/10.1037/0012-1649.39.1.34

Thoman, D. B., White, P. H., Yamawaki, N., & Koishi, H. (2008). Variations of Gender–math Stereotype Content Affect Women's Vulnerability to Stereotype Threat. *Sex Roles*, *58*(9-10), 702–712. http://doi.org/10.1007/s11199-008-9390-x

Twenge, J. M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles*, *36*(5-6), 305–325. http://doi.org/10.1007/BF02766650

Verkuyten, M. (2005). *The social psychology of ethnic identity*. Psychology Press.

Verkuyten, M. (2010). Assimilation ideology and situational well-being among ethnic minority members. *Journal of Experimental Social Psychology*, *46*(2), 269–275. http://doi.org/10.1016/j.jesp.2009.11.007

Vorauer, J. D., Gagnon, A., & Sasaki, S. J. (2009). Salient Intergroup Ideology and Intergroup Interaction. *Psychological Science*, *20*(7), 838–845. http://doi.org/10.1111/j.1467-9280.2009.02369.x

Weisgram, E. S., Bigler, R. S., & Liben, L. S. (2010). Gender, values, and occupational interests among children, adolescents, and adults. *Child Development*, *81*(3), 778–96. http://doi.org/10.1111/j.1467-8624.2010.01433.x

Weisgram, E. S., Dinella, L. M., & Fulcher, M. (2011). The Role of Masculinity/Femininity, Values, and Occupational Value Affordances in Shaping Young Men's and Women's Occupational Choices. *Sex Roles*, *65*(3-4), 243–258. http://doi.org/10.1007/s11199-011-9998-0

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.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. http://doi.org/10.1037/0033-295X.107.1.101

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes Are Attractive Because They Are Easy on the Mind. *Psychological Science*, *17*(9), 799–806. http://doi.org/10.1111/j.1467-9280.2006.01785.x

Wolsko, C., Park, B., & Judd, C. M. (2006). Considering the Tower of Babel: Correlates of Assimilation and Multiculturalism among Ethnic Minority and Majority Groups in the United States. *Social Justice Research*, *19*(3), 277–306. http://doi.org/10.1007/s11211-006-0014-8

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. http://doi.org/10.1037/0022-3514.78.4.635

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. http://doi.org/10.1111/j.1559-1816.1991.tb00535.x

APPENDIX

Appendix A: Study 1 Materials

[Study 1a: Instructions]

First impressions are made very quickly, within only a few seconds, and are often surprisingly accurate. For example, inferences of a person's competence based solely on facial appearance predicted the outcomes of U.S. congressional elections better than chance (e.g., 68.8% of the Senate races in 2004; Todorov et al., 2005). In another study, people were able to detect someone's sexual orientation by looking at their face for just half a second (Rule & Ambady, 2008).

In this study, we are interested in your first impressions of a variety of faces. Specifically, we are interested in your assessment of their:

1. Femininity/Masculinity
2. Likability
3. Attractiveness
4. Age
5. Likelihood of being a scientist
6. Likelihood of being an early childhood educator

You will make these judgments of about 80 faces, half of which are women and half of which are men. There are no right or wrong answers--just go with your gut reactions towards each face.

Depending on your Internet connection, some of the faces may take a moment to load. Please be patient and wait for the face to appear on each page.

[Study 1a: Ratings]

Consider the photo above. To what extent would you say this person is…

Masculine	o	o	o	o	o	o	o	**Feminine**
Likable	o	o	o	o	o	o	o	**Unlikable**
Unattractive	o	o	o	o	o	o	o	**Attractive**

How likely do you think it is that this person is a scientist?

Very unlikely	Unlikely	Somewhat Unlikely	Somewhat Likely	Likely	Very Likely
o	o	o	o	o	o

How likely do you think it is that this person is an early childhood educator?

Very unlikely	Unlikely	Somewhat Unlikely	Somewhat Likely	Likely	Very Likely
o	o	o	o	o	o

What would you guess is this person's age?

25-29 years
30-35 years
36-40 years
41-45 years
45-50 years
51-55 years
56-60 years
61+ years

[Study 1b: Instructions]

Instructions

First impressions are made very quickly, within only a few seconds, and are often surprisingly accurate. For example, inferences of a person's competence based solely on facial appearance predicted the outcomes of U.S. congressional elections better than chance (e.g., 68.8% of the Senate races in 2004; Todorov et al., 2005). In another study, people were able to detect someone's sexual orientation by looking at their face for just half a second (Rule & Ambady, 2008).

In this study, we are interested in your first impressions of a variety of faces. Specifically, we are interested in your assessment of how likely they are to be in three different careers:

1. Likelihood of being a journalist.
2. Likelihood of being a scientist.
3. Likelihood of being an early childhood educator.

You will make these judgments of about 80 faces, half of which are women and half of which are men. There are no right or wrong answers--just go with your gut reactions towards each face.

Depending on your Internet connection, some of the faces may take a moment to load. Please be patient and wait for the face to appear on each page.

[Study 1b: Ratings]

How likely do you think it is that this person is a journalist?

Very unlikely	Unlikely	Somewhat Unlikely	Somewhat Likely	Likely	Very Likely
o	o	o	o	o	o

How likely do you think it is that this person is a scientist?

Very unlikely	Unlikely	Somewhat Unlikely	Somewhat Likely	Likely	Very Likely
o	o	o	o	o	o

How likely do you think it is that this person is an early childhood educator?

Very unlikely	Unlikely	Somewhat Unlikely	Somewhat Likely	Likely	Very Likely
o	o	o	o	o	o

Appendix B. All Traits included in Study 2 Pre-Screen

Positive Words	**Negative Words**
Ambitious	Aloof
Analytical	Arrogant
Capable	Boastful
Caring	Clueless
Compassionate	Cold
Competent	Critical
Decisive	Cutthroat
Driven	Disorganized
Emotional	Ditzy
Encouraging	Egotistical
Gentle	Flakey
Intelligent	Forgetful
Intuitive	Gullible
Kind	Ignorant
Logical	Illogical
Meticulous	Impulsive
Nurturing	Inconsiderate
Objective	Irrational
Patient	Isolated
Rational	Naive
Scientific	Robotic
Skillful	Self-absorbed
Supportive	Selfish
Talkative	Uncertain
Understanding	Uninformed
Warm	Unsociable

Appendix C. Pre-Screen Explicit Measures

[Gender Ideology Items]

We are interested in your personal perspective on gender. Please answer the following questions about how you understand differences and similarities between men and women. There are no right or wrong answers and we are simply interested in your personal opinion.

1. Men and women are naturally suited to different jobs and should continue to do those.

o	o	o	o	o	o	o
Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Agree	Strongly Agree

…all Gender Ideology Items [see Table 2]

[Explicit Gender Stereotypes: Percent Estimate Task]

As you make the following judgments, consider WOMEN in the United States. For each judgment, estimate the percentage, from 0 to 100%, of WOMEN in the US that you believe have each attribute. You should base your judgment on your own opinion or beliefs; there are no right or wrong answers.
In your response, please use only numbers. Do not use the % sign or text. (For example: Correct entry: 50. Incorrect entry: 50%. Incorrect entry: 50 percent)

What percentage of women [men] are incompetent? [ ]
…ditzy?
…Isolated?
…warm?
…naïve?
…logical?
…critical?
…nurturing?
…analytical?
…gentle?
…robotic?
…objective?
…imprecise?
…scientific?
…caring?
…self-absorbed?

[Evaluation: Thermometer Ratings]

There are lots of different groups that live in the United States. We would like to ask you about

your feelings toward a subset of these groups. Consider the scale below. You can think of this as a "feeling" scale that is intended to measure how warmly or how coolly you feel toward a given group. For each of the groups listed below the scale, we would like you to indicate how you feel toward that group. Clearly there are no right or wrong answers. We simply want your feelings toward the groups.

Please use the following scale to answer all of the questions in this section.

1. How warmly or coolly do you feel towards MEN? Please fill in any number from 0 to 100.

2. How warmly or coolly do you feel towards WOMEN? Please fill in any number from 0 to 100.

[Gender-STEM Stereotypes]

Listed below are questions for this section of the prescreen. Please provide a response for every question. If you are given the option to decline to answer a question, then declining to answer is considered a response.

1. According to my own personal beliefs, I generally expect men to do better in math and science than women.

o	o	o	o	o	o	o
1 Strongly Disagree	2	3	4 Neither Agree Nor Disagree	5	6	7 Strongly Agree

2. According to general beliefs in society, men are expected to be better at math and science than women.

o	o	o	o	o	o	o
1 Strongly Disagree	2	3	4 Neither Agree Nor Disagree	5	6	7 Strongly Agree

Appendix D. GNAT Instructions

[Verbal Instructions]

Hi everyone! I'm ________. Please silence your cell phones before we begin.

Welcome to Experiment 1277. Thank you for agreeing to participate in this study. This is a 2-credit experiment and will take approximately 45-50 minutes to complete. If at any time you have any questions, please do not hesitate to ask. To learn more about this study, please read through the informed consent on the computer now.

….

Let me describe this study to you. Throughout this experiment, you will be completing what is called a Go/No-Go Task.

In this task you will be shown a series of words or pictures in the center of the computer screen. Some of these will belong to TARGET CATEGORIES, and others will simply be DISTRACTORS. During the task, if you see a word or picture that belongs to a target category, your job is to *GO*—to press the SPACEBAR as quickly as possible. In contrast, if you see a word or picture that does NOT belong to a target category, you should simply ignore it (*NO-GO*).

For example, let's say that the target category is BIRDS. Your job would then be to *GO*--press the SPACEBAR--every time you see a bird name or image, and to ignore any word or image that is not a bird.

We emphasize that you have to press the SPACEBAR quickly. Each stimulus will only remain on the screen for less than a second. If you make a response after the word or picture has disappeared, the response will be considered incorrect. In addition, if you press the SPACEBAR when the word or picture is NOT in the target category, the response will be incorrect.

To let you know how you are doing, a red "X" will appear if you press the SPACEBAR when you should not. Additionally, a blue "!!" will appear if you were too slow to press the SPACEBAR when you should.

Please note that because the task is so fast, you will make mistakes and that's Okay.

Before we begin the experimental Go/No-Go's, you will complete some practice trials so you can become comfortable with the task and ask me any questions you have about it.

At this point, I will let you each go at your own pace through the practice trials. Are there any questions?

Press RETURN to begin the practice trials.

…

I know this task is challenging and requires a lot of concentration, so I'm going to pass around this candy. Please take some so that you can have it when you need a break.

You are now ready to begin the 12 experimental go/no-go tasks. First, you will have an opportunity to become familiar with all of the stimuli that you will see today. **As a reminder of what they are, there's also a cheat sheet on your desk that you should feel free to refer to when you need it.**

Please take your time and if you need to take a break in between the tasks, feel free to take one. When you finish the Go/No-Go task, you will complete a survey online. Once you are completely done, please wait for me to dismiss everyone at the same time.

Any questions? Go ahead and begin!

…

It looks like all of you are now finished. Thanks so much for your time. Please get your receipt on the way out. If you have any comments or questions, I'm here to talk. Have a nice day!

[“Cheat Sheet”]

Category & Word “Cheat Sheet”

GOOD words:	**BAD words:**
BEACH	COCKROACH
PARADISE	DISEASE
PUPPY	FILTH
SMILE	POISON
SUNRISE	VOMIT
SCIENTIFIC traits:	**UNSCIENTIFIC traits:**
ANALYTICAL	DITZY
LOGICAL	CARELESS
OBJECTIVE	FORGETFUL
METICULOUS	GULLIBLE
SCIENTIFIC	NAÏVE
WARM traits:	**COLD traits:**
CARING	CRITICAL
COMPASSIONATE	ISOLATED
GENTLE	ROBOTIC
NURTURING	SELF-ABSORBED
WARM	UNSOCIABLE

[Self-Paced Computer GNAT Instructions; Pages separated by lines]

Summary of Instructions:

When you see a word or image that belong to target categories, press the SPACEBAR (GO).

When you see a word or image that does not belong to target categories, ignore it (NO-GO).

You must respond quickly. Stimuli will be presented very quickly, appearing for less than a second. This task is very difficult, and you will make mistakes.

Incorrect responses include being too slow to respond (a blue "!!" will appear) and responding to a non-target word or image (a red "X" will appear).

Press RETURN.

To give you an idea of how the task works, you will now complete some practice trials.

Before we begin the practice trials, it is helpful if you are familiar with all of the categories and words we will ever ask you to respond to.

To view all of the categories and words we will ask you to respond to, press RETURN.

During the first practice trial, the target category is BIRDS.

Whenever you see the name [a photo] of a bird, press the SPACEBAR as quickly as possible. Ignore all other things that you see.

When you are ready to begin the task, press RETURN.

Good job! In the next practice trial, the target category is WOMEN.

Whenever you see the name [a photo] of a woman, press the SPACEBAR as quickly as possible. Ignore all other things that you see.

This GO/NO-GO will move slightly faster than the last one, so you will have to respond more quickly.

When you are ready to begin the task, press RETURN.

Good job! In the next practice trial, the target category is MEN.

Whenever you see the name [a photo of] of a man, press the SPACEBAR as quickly as possible. Ignore all other things that you see.

When you are ready to begin the task, press RETURN.

Excellent. Now, the actual experimental GO/NO-GO Tasks will ask you to go to more than one category at a time.

Let's do one last practice. In this practice, GO (press the SPACEBAR) when you see either names of BIRDS or names of CATS. Ignore all other things.

Now that you're warming up, the words will be shown for an even shorter amount of time.

Press RETURN when you are ready to begin.

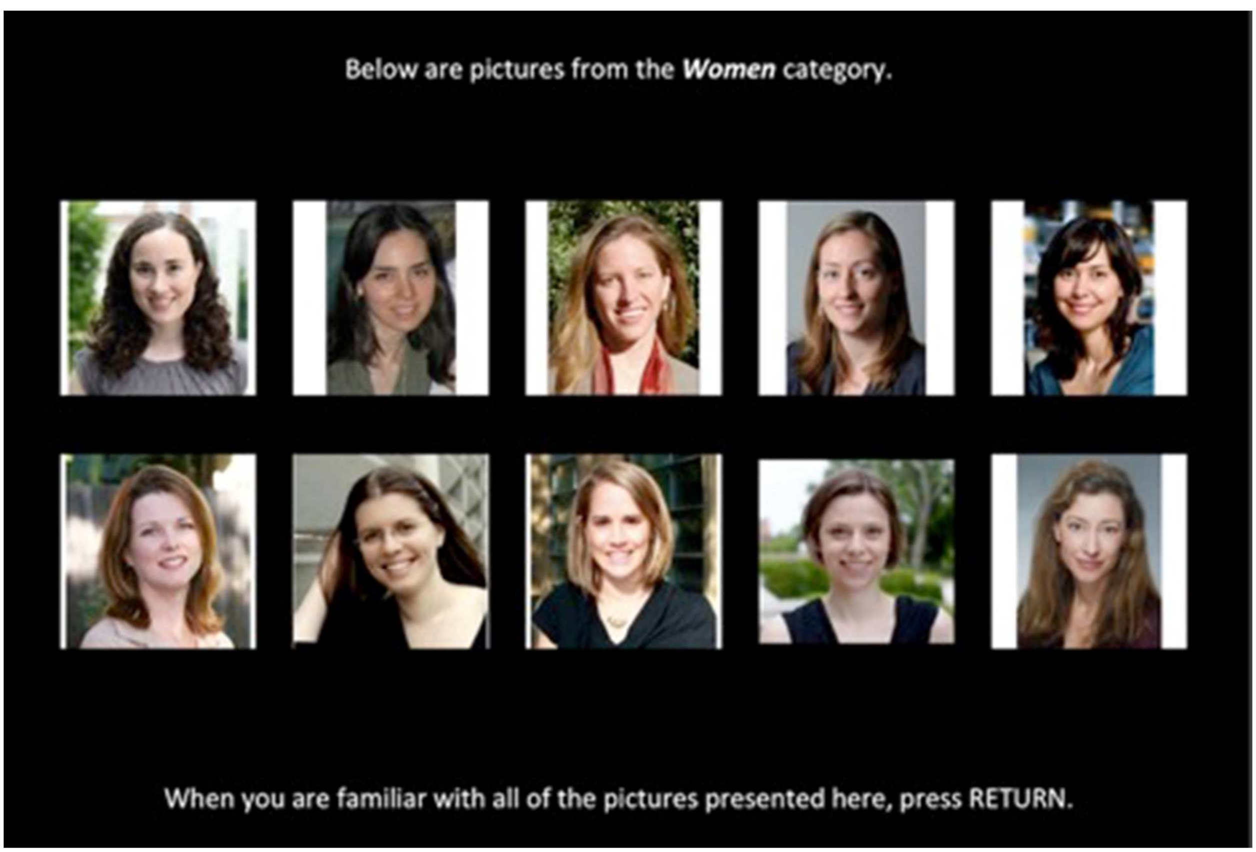
Excellent. Before we begin the actual experimental GO/NO-GO's, it is helpful if you are familiar with all of the categories and stimuli we will ever ask you to respond to.

Throughout this experiment, you will complete 12 different GO/NO-GO tasks that ask you to respond to stimuli from a variety of categories.

In the practice trials, you have already seen the categories MEN, WOMEN, and BIRDS and the stimuli within those categories.

To see the other categories and stimuli we will ask you to respond to, please press RETURN.

[Example Stimuli Slide for Participants in Stereotypic Condition. Participants also saw men.]



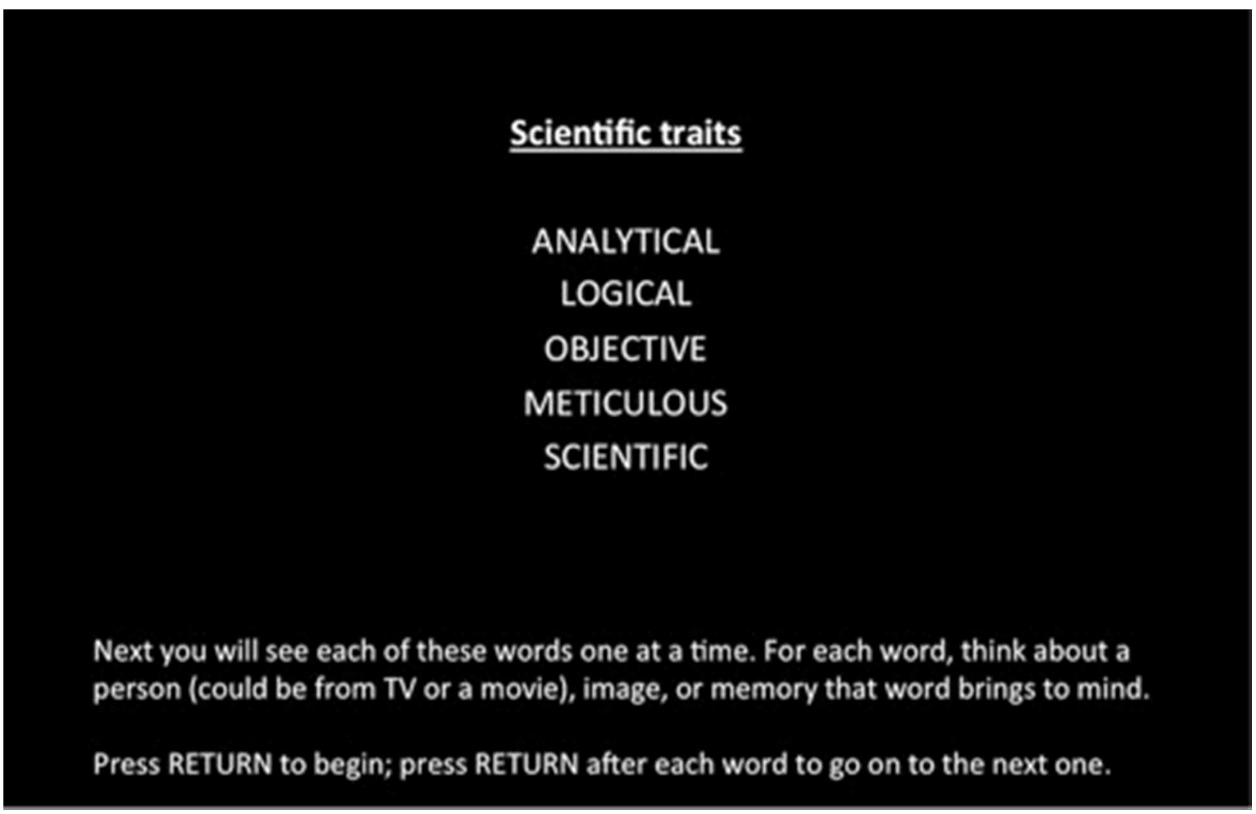
In addition, you will be asked to respond to words from the following categories:

Good Words
Bad Words
Scientific Traits
Unscientific Traits
Warm Traits
Cold Traits

On the following slides, you will see each of these words one at a time. For each word, think about a person, image, or memory that word brings to mind. To advance through the words, press RETURN.

To see all of the words that belong within each category, press RETURN.

[Example Stimuli Slide]



You are now ready to begin the experimental sets of the GO/NO-GO task. These will work the same way as the practice trials, the only difference is that these sets will be longer and there will be more of them.

There will be 12 experimental sets of the GO/NO-GO Task, and each task will include two target categories.

If you feel fatigued and need to take a short break, you may do so between sets. Feel free to have some candy or take a deep breath. However, if you start a set you must finish it without taking a break.

Press RETURN.

If you need a reminder of the words within each category, you can always refer to your word "cheat sheet" (sitting on your desk) to be reminded of the words that belong within a category.

Press RETURN to begin the experimental GO/NO-GO's.

[Example instructions for Women + Scientific Block for a participant in a photo condition]

In this block of trials, your task is to select things from the following categories:

WOMEN or SCIENTIFIC traits

When you see either WOMEN (a photo of a woman) or a SCIENTIFIC trait, you must press the spacebar as quickly as possible while the word is on the screen. You should ignore all other things.

As you have seen, there is very little time to make each decision. To be accurate, it is important that you concentrate. Ask the experimenter if you have any questions.

Press RETURN to continue.

Appendix E. Counterbalancing of GNAT Blocks

Position:	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**	**9th**	**10th**	**11th**	**12th**
Sub1	9	10	6	8	3	12	1	4	5	11	7	2
Sub2	10	7	8	5	9	1	11	6	2	12	4	3
Sub3	6	2	1	11	8	4	10	3	9	5	12	7
Sub4	1	3	10	4	5	2	7	12	11	9	8	6
Sub5	3	1	4	10	2	5	12	7	6	8	9	11
Sub6	5	9	11	6	7	8	4	1	12	3	2	10
Sub7	2	6	3	7	4	11	8	9	1	10	5	12
Sub8	12	11	2	1	10	6	5	8	7	4	3	9
Sub9	4	5	7	3	12	9	2	10	8	6	11	1
Sub10	7	4	5	12	1	3	9	11	10	2	6	8
Sub11	8	12	9	2	11	7	6	5	3	1	10	4
Sub12	11	8	12	9	6	10	3	2	4	7	1	5

Note: These 12 Blocks were presented in one of 12 orders for a given subject. Each subject gets one row, the columns refer to the position number 1st, 2nd etc, and in each cell is the number of the GNAT block presented in that position for that subject.

Appendix F. Laboratory Session Explicit Measures

[Percent Estimate Task]

Consider the group WOMEN [MEN] in the United States. For each judgment, estimate the percentage, from 0 to 100%, of WOMEN [MEN] in the US that you believe have the attribute. You should base your judgment on your own opinion or beliefs. There are no right or wrong answers. Just go with your gut reaction.

What percentage of WOMEN [MEN] are...

	0	10	30	40	50	60	70	80	90	100
Naïve										

[And all other traits from Tables 9 and 10]

[Ambivalent Sexism Inventory (Glick & Fiske, 1996]

Below are a series of statements concerning contemporary society. Please indicate how much you agree or disagree with each statement.

... Men are complete without women.	o Strongly Disagree	o Disagree	o Somewhat Disagree	o Somewhat Agree	o Agree	o Strongly Agree

.... Men should be willing to sacrifice their own well being in order to provide financially for the women in their lives.

.... Most women interpret innocent remarks or acts as being sexist.

.... Once a woman gets a man to commit to her, she usually tries to put him on a tight leash.

.... Women seek to gain power by getting control over men.

.... Every man ought to have a woman whom he adores.

.... Women should be cherished and protected by men.

.... Women, compared to men, tend to have a superior moral sensibility.

.... Women exaggerate problems they have at work.

.... Women are too easily offended.

[Internal/External Motivation to Respond without Sexism; Klonis, Plant, & Devine, 2005]

Below are a series of statements concerning contemporary society. Please indicate how much you agree or disagree with each statement.

	o	o	o	o	o	o	o
... **Being nonsexist toward women is important to my self-concept.**	Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree

Internal Motivation

- According to my personal values, using stereotypes about women is OK (reverse-coded)
- I am personally motivated by my beliefs to be nonsexist toward women.
- Being nonsexist toward women is important to my self-concept.
- Because of my personal values, I believe that using stereotypes about women is wrong.
- I attempt to act in nonsexist ways toward women because it is personally important to me.

External Motivation

- Because of today's PC (politically correct) standards I try to appear nonsexist toward women.
- I try to hide any negative thoughts about women in order to avoid negative reactions from others.
- If I acted sexist toward women, I would be concerned that others would be angry with me.
- I attempt to appear nonsexist toward women in order to avoid disapproval from others.
- I try to act in nonsexist ways because of pressure from others.