Adults and Children Implicitly Associate Brilliance with Men more than Women

Daniel Storage[1]*, Tessa E. S. Charlesworth[2], Mahzarin R. Banaji[2], Andrei Cimpian[3]*

[1]University of Denver, [2]Harvard University, [3]New York University

* Email: Daniel.Storage@du.edu and andrei.cimpian@nyu.edu

In press, *Journal of Experimental Social Psychology*.

Author Note

We are grateful to our participants, including the Urbana-Champaign and New York families who participated. We would also like to thank the research assistants from the Cognitive Development Labs at the University of Illinois and New York University for their help with data collection. The members of the Cognitive Development Lab at New York University provided helpful feedback on a previous draft of this manuscript. This research was supported by National Science Foundation grants BCS-1530669 and BCS-1733897 awarded to Andrei Cimpian and the Dean's Competitive Fund for Promising Scholarship from Harvard University to Mahzarin R. Banaji.

Abstract

Women are underrepresented in careers where success is perceived to depend on high levels of intellectual ability (e.g., brilliance, genius), including those in science and technology. This phenomenon may be due in part to a *gender-brilliance stereotype* that portrays men as more brilliant than women. Here, we offer the first investigation of whether people implicitly associate brilliance with men more than women. Implicit measures are absent from prior research on the gender-brilliance stereotype, despite having the potential to contribute unique information about the prevalence of this stereotype. Across 5 studies (N = 3,618) with 17 Implicit Association Tests using 6 distinct comparison traits (e.g., *creative*, *funny*), we found consistent evidence for an implicit gender-brilliance stereotype favoring men. Indeed, for 5 out of 6 comparison traits (even the male-typed trait *funny*), *male* was associated with *brilliant* and *female* with the comparison trait. Only a physical trait (*strong*) showed a stronger association with *male* than *brilliant* did; none of the psychological traits used as comparisons rivaled *brilliant* in their association to *male*. Evidence for the implicit gender-brilliance stereotype was consistently observed whether the male and female targets were represented with verbal labels or pictures, and whether the pictures depicted White or Black targets. Moreover, the results were robust in both men and women, children and adults, across different regions of the U.S. as well as internationally. This pervasive implicit association of brilliance with men is likely to hold women back in careers perceived to require brilliance.

Keywords: stereotypes, gender, brilliance, Implicit Association Test

Adults and Children Implicitly Associate Brilliance with Men more than Women

Women are underrepresented across a range of careers that are perceived to require high-level intellectual ability (e.g., brilliance, genius), including many science, technology, engineering, and mathematics (STEM) fields but also fields in the social sciences and the humanities, such as philosophy (Cimpian & Leslie, 2017; Leslie, Cimpian, Meyer, & Freeland, 2015; Meyer, Cimpian, & Leslie, 2015). This distinctive pattern of underrepresentation emerges in part because of a stereotype that associates these qualities of genius and brilliance with men more than women (Bian, Leslie, & Cimpian, 2017; 2018; Jaxon, Lei, Shachnai, Chestnut, & Cimpian, 2019; Rivera & Tilcsik, 2019). This stereotype may lead members of fields that value brilliance to perceive women as unsuited for these careers (Bian, Leslie, & Cimpian, 2018) and may perhaps also undermine women's own willingness to pursue careers in these fields (Bian, Leslie, Murphy, & Cimpian, 2018). Our main goal in this paper is to contribute to this growing area of research by conducting the first investigation of the gender-brilliance stereotype using *implicit measures*. Implicit measures have remained entirely absent from prior work on gender-brilliance stereotypes despite having the potential to contribute unique information about the prevalence of this stereotype, as well as its malleability and relation to prejudiced behavior.

We begin by clarifying the content of the stereotype under investigation and differentiating it from other gender stereotypes. We then review the existing evidence that people endorse a gender-brilliance stereotype and, finally, motivate the present set of studies.

The Gender-Brilliance Stereotype

We use the terms *brilliance* and *genius* interchangeably to refer to high-level intellectual ability (Bian, Leslie, Murphy, & Cimpian, 2018; Rattan, Savani, Naidu, & Dweck, 2012). People seem to assume brilliance is unevenly distributed not just across individuals (i.e., some people have it and others do not) but also across *social groups* (i.e., some groups have it and others do not). We will focus on the stereotyped assumption that men are more likely to possess brilliance than women, which we refer to as the *gender-brilliance stereotype*.

The content of the gender-brilliance stereotype differs along two key dimensions from the content of other gender stereotypes that pertain to the intellectual domain. First, unlike gender stereotypes about mathematical, scientific, or verbal ability (e.g., Cvencek, Meltzoff, & Greenwald, 2011; Nosek, Banaji & Greenwald, 2002a), the gender-brilliance stereotype is *general*—it pertains to a quality that cuts across specific domains of intellectual activity; its breadth is likely to make this stereotype particularly powerful. Second, unlike gender stereotypes about competence (e.g., Fiske, Cuddy, Glick, & Xu, 2002; Williams & Best, 1990), the gender-brilliance stereotype pertains to a *particularly high level* of intellectual talent. In principle, a person can think that men and women are equally intelligent on average (i.e., disagree with the gender-competence stereotype) but also think that men are overrepresented among geniuses (i.e., agree with the gender-brilliance stereotype). Consistent with the idea that gender-competence and gender-brilliance stereotypes are empirically distinguishable, a recent meta-analysis of public opinion polls showed a significant decrease from 1946 to 2018 in the prevalence of gender-competence stereotypes favoring men (Eagly, Nater, Miller, Kaufmann, & Sczesny, 2019): The percentage of poll respondents who reported they believed women to be *more* intelligent than men increased steadily over this period, and the percentage of responses indicating that men and women are equally intelligent was generally high (e.g., 85.9% in a nationally-representative 2018 poll). Thus, if we take these data at face value, gender-competence stereotypes favoring men may now be in the past. In contrast, gender-brilliance stereotypes seem to remain widespread, as we review in the next section.

Before describing prior evidence on gender-brilliance stereotypes, we note that these stereotypes are worthy of study not simply because they are different from other gender stereotypes or because they are widespread. These stereotypes are also important to investigate and understand because they seem to hold women back across a wide range of prestigious careers. The more that people in a field believe that these qualities of brilliance and genius are needed for success in their field, the fewer women can be found among its members (e.g., Leslie, Cimpian, et al., 2015; Storage et al., 2016), and this relationship holds even when accounting for fields' reliance on domain-specific skills such as mathematics (Cimpian & Leslie, 2015; Storage et al., 2016). Thus, arriving at a robust, comprehensive understanding of gender-brilliance stereotypes can inform future efforts to increase gender equity in career outcomes.

Previous Evidence on the Gender-Brilliance Stereotype

A major source of evidence for the existence of a gender-brilliance stereotype consists of people's self-reported impressions of familiar others (e.g., their instructors, their children; see Bennett, 1996; Furnham, Reeves, & Budhani, 2002; Kirkcaldy, Noack, Furnham, & Siefen, 2007; Rivera & Tilcsik, 2019; Storage et al., 2016; Tiedemann, 2000). For example, students use the words *brilliant* and *genius* more often to describe male instructors than female instructors in their reviews on RateMyProfessors.com (Storage et al., 2016; see also Schmader, Whitehead, & Wysocki, 2007). Similarly, participants refer more male than female acquaintances for jobs said to require high levels of intellectual ability but not for jobs said to require high levels of motivation (Bian, Leslie, & Cimpian, 2018).

Consistent with this evidence, other studies have found a gender-brilliance stereotype favoring men when assessing participants' explicit judgments of unfamiliar women and men (Bian, Leslie, & Cimpian, 2017; 2018; see also Raty & Snellman, 1997; Rivera & Tilcsik, 2019).

For example, when shown pairs of unfamiliar women and men and asked to guess which person in each pair was “really, really smart,” children as young as 6 years of age from different regions of the U.S. chose significantly more men than women (Bian et al., 2017; Jaxon, Lei, et al., 2019). Similarly, when children were shown a set of unfamiliar peers and asked to select several teammates for a game described as being for children who “are really, really smart,” they first selected teammates of their own gender but then proceeded to select mostly boys as teammates (Bian, Leslie, & Cimpian, 2018). This difference was not observed when the same game was described as being for children who “try really, really hard,” suggesting that the preference for boys as teammates was specific to the “brilliance” game. Note that even though these judgments concerned individuals (rather than groups), the individuals were *unfamiliar* to participants, which means that their responses likely reflected their general concepts of males and females.

Aim of the Present Research

The main aim of the present research was to extend prior work on the gender-brilliance stereotype by investigating this stereotype with *implicit measures*. In social cognition, a distinction has long been drawn between explicit and implicit measures of stereotypes (e.g., Greenwald & Banaji, 1995). Examining both explicit and implicit measures is important for a complete understanding of the prevalence of a stereotype, its implications for behavior, and its malleability to interventions. That is, explicit and implicit stereotypes have been shown to differ in theoretically important ways, including (1) their prevalence across different demographic groups, with implicit attitudes being more consistent across demographic groups (e.g., both old and young people prefer young people on implicit measures; Nosek et al., 2007); (2) their ability to predict behaviors, with implicit stereotypes being more likely to predict spontaneous behaviors (Kurdi, Seitchik, et al., 2019); (3) the mechanisms by which they are most readily changed, with implicit attitudes/stereotypes argued to change more through associative processes and explicit attitudes/stereotypes through propositional processes (Gawronski & Bodenhausen, 2006; cf. Kurdi & Banaji, 2017), and (4) their responsiveness to social changes, with implicit attitudes/stereotypes often changing relatively more slowly than explicit attitudes/stereotypes (Charlesworth & Banaji, 2019a). Because implicit and explicit measures provide complementary perspectives on a stereotype, the main goal of the present work was to test whether the gender-brilliance stereotype is also present on implicit measures of stereotyping.

We employed a common measure of implicit cognition, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), which indirectly assesses the degree of overlap between concepts (e.g., *brilliant* and *male*, *creative* and *female*) via response latencies in a sorting task. Because this task asks participants to respond as quickly and accurately as possible and is therefore based on participants' automatic responses, the IAT can identify stereotypes that participants may be unable or unwilling to report (Greenwald & Banaji, 1995).

Although the validity and reliability of the IAT have recently been challenged (e.g., Schimmack, 2019; Tetlock & Mitchell, 2009; for responses to these types of concerns, see Gawronski, 2019; Jost, 2019; Vianello & Bar-Anan, 2020), the IAT nevertheless offers distinct information about a stereotype that complements the information obtained with explicit measures. In addition, the IAT has been used extensively to study stereotypes (e.g., Kurdi, Seitchik, et al., 2019; Nosek et al., 2007), and the evidence to date speaks to its construct and predictive validity as a measure of stereotypic associations (for a review, see Nosek, Greenwald, & Banaji, 2007). For instance, Nosek et al. (2009) found that nations with stronger implicit gender-math stereotypes favoring males also showed a larger male advantage in 8th grade standardized math performance relative to countries with weaker gender-math stereotypes. It is thus likely that an implicit gender-brilliance stereotype, if present, has implications for behavior both performed by, and directed toward, women in fields and careers where brilliance is valued (e.g., physics, philosophy). Our goal here is to provide the first investigation of this implicit stereotype.

Overview of Studies

In the current research, we tested implicit gender-brilliance associations across five studies with 3,618 child and adult participants from different regions of the U.S., as well as from 78 other countries. In a subset of these studies, we also collected data on participants' explicit gender-brilliance associations, which enabled direct comparisons of the strength of explicit and implicit stereotypes.

Study 1 provided a first test of the implicit gender-brilliance stereotype by measuring the association of *male* (vs. *female*) with *brilliant* vs. a comparison trait (either *creative* or *happy*) in a geographically diverse sample of U.S. participants.

Study 2 examined the *relative strength* of the association between *male* (vs. *female*) and *brilliant* relative to four comparison traits, selected to be both female-typed (*friendly, beautiful*) and male-typed (*strong, funny*). This approach can therefore position the strength of the *male-brilliant* association relative to other gender-trait associations. That is, will the *male-brilliant* association be overridden when the comparison trait is male-typed, indicating that the *male-brilliant* association may be weaker than *male-funny* or *male-strong*? Or will participants still associate *brilliant* with *male* even when this requires associating a male-typed trait with *female*, thus indicating that *male-brilliant* may be stronger than *male-funny* or *male-strong*?

Studies 3–5 built outward to examine the generalizability of the implicit gender-brilliance stereotype across stimuli, participant ages, and geography. Specifically, Study 3 explored the intersectionality of the implicit gender-brilliance stereotype by testing whether the stereotype persists when the gender stimuli consist of Black females and males (rather than White females and males, as in Study 1). Study 4 investigated whether the implicit stereotype is present early in life, by ages 9 and 10, in line with prior developmental work using explicit measures (e.g., Bian et al., 2017). Finally, Study 5 provided an initial investigation of the cross-national generalizability of the implicit gender-brilliance stereotype.

General Method

Across all studies, we administered an IAT probing whether participants associate brilliance with men more than women. Participants' reaction times in the IAT were measured as they sorted (1) stimuli related to the category *male*, (2) stimuli related to the category *female*, (3) words related to the trait *brilliant* (e.g., genius, brilliant, super-smart), and (4) words related to a similarly-positive comparison trait (see Table 1). Reaction times were then compared across two critical blocks: (1) a "stereotype-congruent" block, in which *male* and *brilliant* were assigned the same response key (and thus sorted together), while *female* and the comparison trait were assigned a different response key (and sorted together); and (2) a "stereotype-incongruent" block, in which the pairings were reversed (*male* + *comparison trait* and *female* + *brilliant*). The order of the critical blocks was randomized across participants.

The logic of the IAT is as follows: If *brilliant* is indeed more associated with *male* than *female* in people's minds, then participants will be faster when *brilliant*- and *male*-related stimuli are sorted with the same key (i.e., the "stereotype-congruent" block) than when *brilliant*- and *female*-related stimuli are sorted with the same key (i.e., the "stereotype-incongruent" block). The difference between a participant's reaction times on congruent vs. incongruent blocks, expressed as a fraction of the overall variability in the participant's responses, is known as a *D* score and provides an index of the extent to which the participant more readily associates brilliance with men and the comparison trait with women. Across all studies reported here, positive *D* scores indicate a *male-brilliant*/*female-comparison trait* association, and negative *D* scores indicate a *female-brilliant*/*male-comparison trait* association. We refer readers to Greenwald, McGhee, and Shwartz (1998), Greenwald, Nosek, and Banaji (2003), and Nosek, Greenwald, and Banaji (2005) for more information on how the IAT is structured and scored.

Comparison Traits. Across the five studies, we used six unique comparison traits (*creative*, *happy*, *strong*, *funny*, *friendly*, and *beautiful*; see Table 1) to rule out the possibility that participants' responses were driven primarily by an association between *female* and a particular comparison trait (for a similar strategy, see Jajodia & Earleywine, 2003; Sherman et al., 2003). In addition to ruling out this alternative, the inclusion of six unique comparison traits allows us to situate the strength of the *male-brilliant* association relative to other *male-trait* associations (such as *male-funny* or *male-strong*). That is, using six comparison traits can answer a key question: At what point, if ever, does a *male-brilliant* association disappear? Is there an identifiable comparison trait that can override this association, or is it present regardless of the comparison?

An alternative strategy for measuring gender-brilliance stereotypes would have been to use a single-category IAT (SC-IAT) comparing the associations of *male* vs. *female* with a single trait (*brilliant*). We opted for the strategy of using a range of comparison traits in a standard IAT (as described above) for three reasons. First, the psychometric properties of the SC-IAT are weak, with lower predictive and internal validity than the standard IAT (Kurdi, Seitchik, et al., 2019). Second, in the SC-IAT, participants may adopt simplifying strategies that make it difficult to interpret whether conceptual *associations* are actually being measured (Nosek, Greenwald, &

Banaji, 2007). For instance, if *male* and *brilliant* are assigned the same response key and *female* is assigned the other key, participants can categorize stimuli simply based on whether they match or mismatch *female* rather than based on an association with *brilliant*. Third, as mentioned above, our chosen approach of using multiple comparison traits allows us to situate the *male-brilliant* association relative to other stereotypic associations. It is only by using the standard IAT with multiple comparison traits that we are able to assess which traits, if any, are more associated with men (or less associated with women) than the trait *brilliant.*

Table 1. *The IAT Stimuli in Studies 1–5*

Study	Female Stimuli	Male Stimuli	Brilliance Stimuli	Comparison Trait Stimuli
1, 4, 5	8 photographs of White women (*female*)	8 photographs of White men (*male*)	*genius*, brilliant, super-smart	*creative*, artistic, super-imaginative
1	8 photographs of White women (*female*)	8 photographs of White men (*male*)	*genius*, brilliant, super-smart	*happy*, joyful, super-upbeat
2a, 2b	*female*, she, her, woman, women	*male*, he, him, man, men	*super smart*, brilliant, genius, brainiac	*very beautiful*, elegant, graceful, pretty
2a, 2b	*female*, she, her, woman, women	*male*, he, him, man, men	*super smart*, brilliant, genius, brainiac	*very friendly*, outgoing, kindly, chatty
2a, 2b	*female*, she, her, woman, women	*male*, he, him, man, men	*super smart*, brilliant, genius, brainiac	*very funny*, entertaining, hilarious, witty
2a, 2b	*female*, she, her, woman, women	*male*, he, him, man, men	*super smart*, brilliant, genius, brainiac	*very strong*, brave, bold, tough
3	8 photographs of Black women (*female*)	8 photographs of Black men (*male*)	*genius*, brilliant, super-smart	*creative*, artistic, super-imaginative

Note. Italicized stimuli represent the trait/category label. The stimuli within each category were included based on the authors' judgment that they best represented the category. The pictures of females and males in Studies 1, 3, 4, and 5 were taken from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) and were matched in terms of attractiveness, age, and positive emotionality (for details, see Appendix S1 in the Supplementary Online Materials [SOM]).

Format, Scoring, and Administration. All IATs followed a standard 7-block format (Greenwald, McGhee, & Shwartz, 1998), and scores were analyzed using the publicly available R packages *IATScore* (Storage, 2017) and *IATanalytics* (Storage, 2018), in accordance with the updated scoring algorithm of Greenwald et al. (2003). Adults completed the IAT online, whereas children (Study 4) completed it on a computer in a lab or at their school, in the presence of a researcher.

Explicit Measures. In a subset of studies, we administered several explicit measures at the end of the sessions. First, we included explicit measures of the gender-brilliance stereotype, which allowed us to compare participants' explicit endorsement of the gender-brilliance stereotype with the strength of their implicit associations. We administered two different types of explicit measures, some akin to traditional scales (see Study 1) and others more similar to the comparative format of the IAT, in that they asked participants to self-report whether they associated the trait *brilliant* with men more than women or vice-versa (see Study 2). In both cases, we compared participants' explicit endorsement of the gender-brilliance stereotype with their IAT scores.

Second, we included several measures of gender prejudice and political orientation, which allowed us to explore the correlates of participants' gender-brilliance IAT scores. If the IAT captures the implicit belief that men (more than women) are brilliant—rather than merely the implicit belief that women (more than men) possess some positive comparison trait—then scores on the IAT should correlate positively with explicit measures of prejudice against women (e.g., Old-Fashioned Sexism; Swim, Aikin, Hall, & Hunter, 1995), as well as with measures of political conservatism, which is often accompanied by negative views of lower-status groups (e.g., Altemeyer, 1981; Christopher & Mull, 2006).

For brevity and because implicit-explicit comparisons are not the primary focus of the present research, these comparisons are reported only for Studies 1 and 2, where they bear on our research questions; all other analyses on this topic are reported in the Supplementary Online Materials (SOM; see Tables S2 and S3) and on the Open Science Framework (OSF): https://osf.io/8xzek/?view_only=acea685941d6443e8bedd00b71cbe961.

Power Analyses. We used G*Power 3.1 to perform sensitivity power analyses for all five studies (Faul, Erdfelder, Lang, & Buchner, 2007). In particular, we calculated the power of the one-sample *t* tests of participants' *D* scores against 0, since *D* scores significantly above 0 provide evidence of an implicit stereotype associating *male* with *brilliant* and *female* with a comparison trait. For all analyses, the power level was set at 80% and the alpha level at .05 (two-tailed tests). These analyses revealed that all of our studies were adequately powered to detect small effects (Cohen's *d*s ≥ 0.28). The study with the lowest power was Study 4 (N = 103), which was still powered to detect effects as small as d = 0.28.

Study 1

Study 1 provides the first test of participants' implicit gender-brilliance stereotype. Specifically, we tested whether participants associate the trait *genius* (genius, brilliant, super-smart) with the category *male* more than the category *female*. Because the IAT is a relative measure and is therefore equally dependent on the choice of the critical trait (*genius*) and the comparison trait, we began with two semantically distinct comparison traits, tested in separate IATs: *creative* (creative, artistic, super-imaginative) and *happy* (happy, joyful, super-upbeat). Although additional comparison traits are tested in Study 2, including specifically male-typed traits (i.e., *funny, strong*), the comparison traits of *creative* and *happy* were intentionally chosen because (1) they are not consistently associated with either females or males (e.g., Bem, 1974;

Garg, Schiebinger, Jurafsky, & Zou, 2018; Helmreich, Spence, & Wilhelm, 1981) and (2) are similar to *genius* in positivity and desirability, as established by norming data on an independent sample (see Appendix S2 in the SOM). To increase the generalizability of our test of implicit gender-brilliance stereotypes among U.S. adults, we recruited participants from three distinct sources: Mechanical Turk (MTurk) and the subject pools of a public and private university from different regions of the country (University of Illinois at Urbana-Champaign [UIUC] and New York University [NYU], respectively).

Participants

Participants (N = 818; 520 women, 297 men, 1 other) were recruited from MTurk (n = 264, M_{age} = 34.0 years, range = 18–69 years), UIUC (n = 276, M_{age} = 19.5 years, range = 18–24 years), and NYU (n = 278, M_{age} = 19.6 years, range = 17–40 years). An additional 53 participants (34 from MTurk, 15 from UIUC, 4 from NYU) were excluded from the final sample because of criteria related to IAT scoring (e.g., if they responded faster than 300 milliseconds on more than 10% of trials, which likely suggests that they clicked quickly through the task; Greenwald et al., 2003), missed attention checks, or because they had IP addresses from outside the U.S. The final sample was 48.8% White, 7.4% Black, 27.9% Asian, 9.3% Hispanic, and 6.6% other. Participants received $1.75 (MTurk) or course credit (UIUC, NYU) for their participation.

The sample size was determined a priori such that each of the two IATs (with *creative* and *happy* as a comparison trait, respectively) would be tested on at least 125 participants in each of the three recruitment venues ($2 \times 3 \times 125 = 750$). We oversampled by approximately 15% relative to the target sample size to allow for exclusions. The same procedure was used to determine the sample size of Study 3, which was similar to the present study in that it also used *creative* as a comparison trait and included participants from UIUC and NYU. No participants were added after initial data analysis.

Procedure and Materials. In this and all subsequent studies, we report all measures and exclusions. In Study 1, participants completed one of two IATs that differed only in the comparison trait: *creative* (n = 413) or *happy* (n = 405; see Table 1 for stimuli). After completing the IAT, the participants completed a battery of explicit measures (see Table 2). These measures included the Gender-Brilliance Stereotype Endorsement Scale, an explicit measure of the extent to which participants endorse the gender-brilliance stereotype (e.g., "One is more likely to find a male with a genius-level IQ than a female with a genius-level IQ"; 1 = *strongly disagree* to 9 = *strongly agree*; Bian, Leslie, & Cimpian, 2018). Participants also filled out the Gender-Brilliance Stereotype Awareness Scale, on which they rated the extent to which people in *society* more generally—rather than the participants personally—endorse the items in the preceding (Endorsement) scale. This is a measure of participants' *awareness* or *knowledge* of the gender-brilliance stereotype, which is distinct from personal endorsement (Ashmore & Del Boca, 1981; Devine, 1989). The rest of the measures consisted of various scales of sexism and political conservatism (see Table 2). We expected scores on the IAT to correlate positively with these scales, which would support the claim that the IAT captures a *male-brilliant* association rather just positive stereotypes about women (i.e., that they are creative and happy).

Table 2. *Explicit Measures of Gender Attitudes and Political Orientation in Study 1*

Measure	Targeted Construct (and Key Reference)
Gender Attitudes and Prejudice	
Gender-Brilliance Stereotype Endorsement and Awareness Scale	Explicit *endorsement* of the stereotype that men are more brilliant than women, as well as *awareness* of the extent to which society endorses this stereotype (Bian et al., 2018)

Old-Fashioned Sexism Scale	Endorsement of traditional gender roles and beliefs (Swim et al., 1995)
Modern Sexism Scale	Forms of sexism that are more common in today's society (e.g., denial of ongoing discrimination toward women) (Swim et al., 1995)
Liberal Feminist Attitude and Ideology Scale[a]	Gender role attitudes, goals of feminism, and feminist ideology (Morgan, 1996)
Political Orientation	
Right-wing Authoritarianism Scale	Deference to established authorities, aggression toward out-groups, support for traditional values (Altemeyer, 1981)
Conservatism Item	Political liberalism vs. conservatism (1 item)

[a]To avoid participant fatigue, we included only the "Global Goals" and "Discrimination and Subordination" subscales.

Results and Discussion

Is There an Implicit Gender-Brilliance Stereotype? The mean IAT *D* score across all participants was 0.22 [0.20, 0.24], which was significantly different from 0, $t(817) = 18.79$, $p < .001$, $d = 0.66$, indicating a reliable *male-genius*/*female-comparison trait* association.[1] 76% of participants showed *D* scores above 0. The prevalence of this stereotype is similar to that of other widely-tested implicit gender stereotypes. For instance, in prior work 72% and 76% of participants showed *D* scores above 0 on IATs measuring gender-science and gender-career stereotypes, respectively (Nosek et al., 2007).

The implicit gender-brilliance stereotype was present both when the comparison trait was *creative*, $D = 0.24$ [0.21, 0.27], $t(412) = 14.70$, $p < .001$, and when the comparison trait was *happy*, $D = 0.19$ [0.16, 0.22], $t(404) = 11.90$, $p < .001$. Moreover, the stereotype was present in

[1] Stereotyped associations appear stronger when the congruent trials make up Blocks 3 and 4 of the IAT rather than Blocks 6 and 7 (e.g., Greenwald et al., 1998; Nosek, Greenwald, & Banaji, 2002b). We found this order effect here as well, $D_{CongruentFirst} = 0.32$ [0.28, 0.35] vs. $D_{IncongruentFirst} = 0.13$ [0.10, 0.16], $t(816) = 8.45$, $p < .001$.

both women (*D* = 0.20 [0.17, 0.23]) and men (*D* = 0.25 [0.21, 0.29]), *t*s > 12.77, *p*s < .001, although it was overall slightly stronger in men, $t(815) = 1.99$, $p = .047$, $d = 0.14$. Closer inspection revealed that men's stereotypes were only stronger when *creative* was the comparison trait, $t(411) = 4.87$, $p < .001$, $d = 0.48$; in contrast, women's stereotypes were stronger when *happy* was the comparison trait, $t(402) = 2.20$, $p = .028$, $d = 0.22$. This difference was not predicted and is not easy to interpret, so we do not speculate about its source. The key take-away is that both IATs, regardless of whether *creative* or *happy* was the comparison trait, showed a *male-genius* association among both women and men (all *p*s < .001). Finally, separate analyses within each sample (MTurk, UIUC, NYU) found a significant gender-brilliance stereotype for both comparison traits and for both women and men, regardless of the sample (see OSF). These results suggest that the implicit gender-brilliance stereotype is pervasive among U.S. adults.

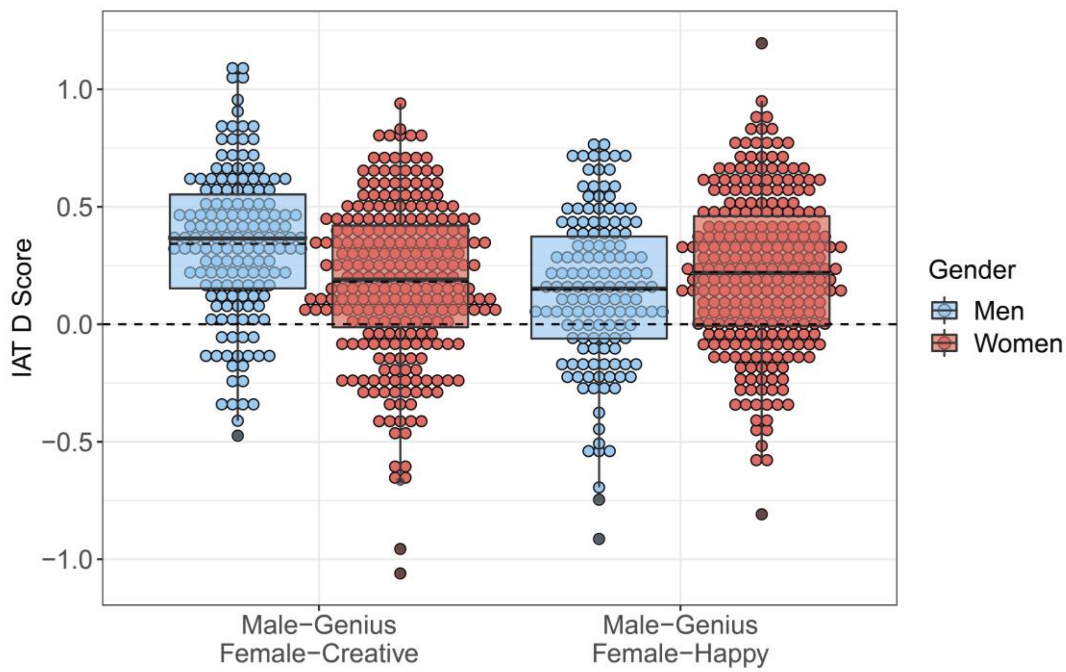

Figure 1. Dot plots (with box plot overlays) of participants' IAT *D* scores in Study 1, by comparison trait (*creative* vs. *happy*) and participant gender. Each dot represents a single participant's *D* score. Solid line = median; dashed line = mean.

Is There an Explicit Gender-Brilliance Stereotype? To assess whether participants explicitly endorse a gender-brilliance stereotype, we compared their responses on the Gender-Brilliance Endorsement Scale against the midpoint (namely, 5). The results revealed that participants did not, in fact, endorse a gender-brilliance stereotype on this explicit measure: Their average agreement levels (M = 3.30 [3.19, 3.41]) were significantly *below* the scale midpoint, $t(817) = -30.1$, $p < .001$, $d = -1.05$. This result contrasts not just with participants' scores on the IAT but also with the studies reviewed in the introduction (e.g., Storage et al., 2016), which found evidence of a gender-brilliance stereotype in people's explicit judgments.

What might explain this contrast? To speculate, the present explicit measure is considerably more *direct* than those used in previous research. None of the studies described above asked participants point-blank if they believed that men are more brilliant than women. Perhaps many participants are simply unaware of holding this belief: Although they *do* believe at some level that men are more brilliant and use this belief to guide explicit judgments and descriptions (e.g., Storage et al., 2016), participants may be unable to introspect on holding this belief. Another possibility, complementary to the first, is participants are unwilling to report this belief: The directness of the items may have prompted participants to be concerned about appearing biased, which may have led them to strategically underreport their endorsement of the gender-brilliance stereotype.

Notably, participants reported much higher *awareness* or *knowledge* of the gender-brilliance stereotype, measured here via participants' reports of the extent to which they believe society endorses this stereotype (M = 5.80 [5.68, 5.92] on the same 1–9 scale). In fact, participants' reports of societal endorsement were significantly above the scale midpoint, $t(817) = 13.0$, $p < .001$, $d = 0.45$. Thus, participants reported that *others*—but not they themselves—

think of brilliance and genius as male qualities, a pattern that suggests this explicit stereotype may in fact be prevalent in the general population (in line with the evidence reviewed above; e.g., Bian, Leslie, & Cimpian, 2018; Rivera & Tilcsik, 2019; Storage et al., 2016).[2]

Relationship Between Implicit and Explicit Gender-Brilliance Stereotypes. Participants' IAT scores were positively correlated with their explicit endorsement of the gender-brilliance stereotype, $r(816) = .08$, $p = .026$, but not with their explicit awareness of it, $r(816) = -.02$, $p = .60$. The positive correlation between the gender-brilliance IAT and participants' explicit endorsement of this stereotype provides some evidence for the validity of this IAT as a measure of the *male-brilliant* association. In addition, the small magnitude of this correlation is consistent with prior evidence for the distinctiveness of implicit and explicit measures of stereotyping (e.g., Nosek et al., 2007). The absence of a correlation between participants' IAT scores and their awareness of the gender-brilliance stereotype in their society may suggest that the gender-brilliance IAT is predominantly measuring individual-level stereotypes (e.g., Banaji, Nosek, & Greenwald, 2004) rather than culture- or society-level associations (Payne, Vuletich, & Lundberg, 2017), although it is also possible that individual self-reports of stereotype awareness do not capture society-level beliefs with sufficient precision.

Relationship Between Implicit Gender-Brilliance Stereotypes and Measures of Gender Bias and Political Conservatism. Participants with higher IAT scores scored significantly higher in Old-Fashioned Sexism, Modern Sexism, Right-Wing Authoritarianism, and political conservatism, rs > .07, ps < .027, and (also as expected) significantly lower on the

[2] Because some of the items in the Endorsement and Awareness scales pertained to broader gender differences in intellectual ability rather than brilliance per se, we also analyzed just the two items that capture the explicit gender-brilliance stereotype most directly: "One is more likely to find a male with a genius-level IQ than a female with a genius-level IQ" and "Extreme intellectual brilliance is more common in men than in women." For these items as well, explicit endorsement of the gender-brilliance stereotype was significantly below the scale midpoint ($M = 2.74$ [2.60, 2.88], $t(816) = -32.2$, $p < .001$, $d = -1.13$) and awareness was significantly above ($M = 5.62$ [5.48, 5.77], $t(816) = 8.42$, $p < .001$, $d = 0.29$).

Feminist Ideology Scale, $r(816) = -.10$, $p = .004$. All of these significant correlations are in line with the interpretation that the gender-brilliance IAT captures *negative* views about women (i.e., that they *are not* brilliant) rather than exclusively positive stereotypes (i.e., that women *are* creative).

Conclusion. The results from Study 1 demonstrate the first clear evidence for a pervasive implicit gender-brilliance stereotype among U.S. adults. These results also highlight the benefit of using implicit measures as a complement to explicit measures, which in this study provided a mixed picture (i.e., no evidence of explicit endorsement of gender-brilliance stereotypes but evidence of awareness of these stereotypes).

Study 2

Our main goal in Study 2 was to estimate the strength of the implicit gender-brilliance stereotype with greater precision by using four comparison traits. The comparison traits were chosen to vary in the extent to which they are typically (explicitly) associated with males vs. females, as indicated by our norming data (see Appendix S3 in the SOM). Specifically, we included a strongly male-typed trait (*strong*), a moderately male-typed trait (*funny*), a strongly female-typed trait (*beautiful*), and a moderately female-typed trait (*friendly*). Using these traits, we can triangulate on the relative strength of the *male-brilliant* association. For instance, if the association between *male* and *brilliant* is stronger than the association between *male* and *funny*, we should observe positive *D* scores suggesting a *male-brilliant* and *female-funny* association; conversely, if the association between *male* and *brilliant* is weaker than that between *male* and *funny*, we should observe negative *D* scores suggesting a *male-funny* and *female-brilliant* association. We investigated these questions in an initial sample of participants from Project Implicit (Study 2a), as well as a preregistered replication sample recruited from the same participant population (Study 2b).[3]

Study 2b also addressed a potential alternative explanation for the implicit gender-brilliance stereotype observed in Study 1. Implicit attitudes are often more positive toward females than males (Dunham, Baron, & Banaji, 2016). If participants' implicit attitudes happen to also be more positive toward the comparison traits (e.g., *beautiful*) than the target trait, then participants might be faster on the "stereotype-congruent" blocks of the IAT simply because positively-valenced *female* is grouped with a positively-valenced comparison trait, forming a coherent pairing (Kurdi, Mann, Charlesworth, & Banaji, 2019). If this alternative explanation is correct, then the results of the IATs might not reflect true semantic or meaning-based stereotypic associations of the sort we set out to investigate. We addressed this possibility by concurrently measuring (in Study 2b only) respondents' implicit stereotypes and implicit valenced attitudes toward the target and comparison traits and examining whether the implicit gender-brilliance stereotype is still present after accounting for the variance explained by the valence of participants' implicit associations with the target and comparison traits.

Participants

Both the initial sample (Study 2a; N = 760; 485 women, 270 men, 5 other; M_{age} = 38.5 years, range = 18–78 years) and the replication sample (Study 2b; N = 1,201; 624 women, 577 men; M_{age} = 34.2 years, range = 16–88 years) consisted of volunteer respondents recruited through the Project Implicit research website (https://implicit.harvard.edu/implicit/). An additional 41 and 309 participants were excluded from Study 2a and Study 2b, respectively, because they failed to complete the IAT. The sample for Study 2a was 75.0% White, 7.5% Black, 3.0% Asian, and 9.4% other; the sample for Study 2b was 79.5% White, 9.1% Black,

[3] Studies 1, 3, 4, and 5 were collected more than a year before Study 2, as part of a dissertation and before preregistration was common practice. This is why only Study 2 was preregistered.

3.2% Asian, and 11.3% other.

The size of the sample for Study 2a was determined a priori so that each of the four IATs (which differed in the comparison trait used) would be tested on approximately 200 participants. We increased the sample size per IAT relative to Study 1 (200 vs. 125) because one of the goals of Study 2 was precise estimation of (potentially smaller) effect sizes. We did not, however, make an allowance for exclusions in this study. Thus, the 41 participants who were excluded from Study 2a came out of the allotment of 800 participants that we received from Project Implicit. The size of the sample for Study 2b was preregistered based on a priori power analyses (https://aspredicted.org/ss36j.pdf). No participants were added after initial data analysis.

Materials and Procedure

The IATs. In both Study 2a and Study 2b, participants were randomly assigned to complete one of four IATs comparing the trait *super smart* with one of four comparison traits (*very strong*, *very funny*, *very friendly*, or *very beautiful*), chosen to represent the range from strongly masculine to strongly feminine attributes (see Appendix S3 in the SOM for norming data). We switched to trait labels that use intensifiers (namely, *super* and *very*) to better equate the positivity and desirability of the brilliance-related trait label and the comparison trait label (see the norming data in Appendix S3). Another difference from Study 1 was that the gender stimuli consisted of gender-related words (e.g., "woman," "man"; see Table 1 for full list of IAT stimuli) rather than pictures of females and males. These minor stimulus changes provided an opportunity to test the robustness of participants' implicit gender-brilliance stereotypes.

Participants in Study 2b also completed an attitude IAT assessing implicit positivity/negativity towards the two traits (e.g., *super smart* and *very beautiful*). In the attitude IAT, the trait stimuli were identical to those in the stereotype IAT, but the gender categories were replaced with *good* (good, best, amazing, excellent, wonderful) and *bad* (bad, awful, worst, terrible, horrible). The order of the attitude and stereotype IATs was counterbalanced across participants.

The Explicit Stereotype Measure. To enable a better-matched comparison between implicit and explicit gender-brilliance stereotypes, we administered an explicit stereotype measure that was more analogous to the IAT, in that it assessed *relative associations between concepts* (rather than endorsement of scale items, as in Study 1). After completing the IAT(s), all participants self-reported the extent to which they associate *super smart* with males vs. females ("To what extent do you associate the quality of SUPER SMART with men and women?"; 1 = *strongly associate with men*, 2 = *moderately associate with men*, 3 = *slightly associate with men*, 4 = *equally associate with both men and women*, 5 = *slightly associate with women*, 6 = *moderately associate with women*, 7 = *strongly associate with women*). Participants also reported the corresponding association for their assigned comparison trait (i.e., *very beautiful*, *very friendly*, *very funny*, or *very strong*). Stereotype awareness was not measured in this study.

Variance Decomposition Analysis

A variance decomposition analysis was performed to examine the amount of variance overlap between implicit stereotypes and implicit valenced attitudes toward the traits (see Kurdi, Mann, et al., 2019, as well as our preregistration: https://aspredicted.org/ss36j.pdf). Specifically, variance for each stereotype IAT was decomposed into (1) error variance, (2) true variance accounted for by the attitude IAT assessing implicit positivity/negativity towards the target and comparison traits ("overlapping variance"), and (3) residual true variance ("independent variance").[4] Significance tests for each variance component were performed using bootstrapped

[4] Error variance was calculated as 1 – the internal consistency of the stereotype IAT (calculated as the mean of 1,000 split half correlations). The variance accounted for by the attribute attitude was calculated as the disattenuated

confidence intervals. If the independent variance is greater than zero, then it can be inferred that the measured implicit stereotype cannot be accounted for exclusively by artifactual valence-based associations.

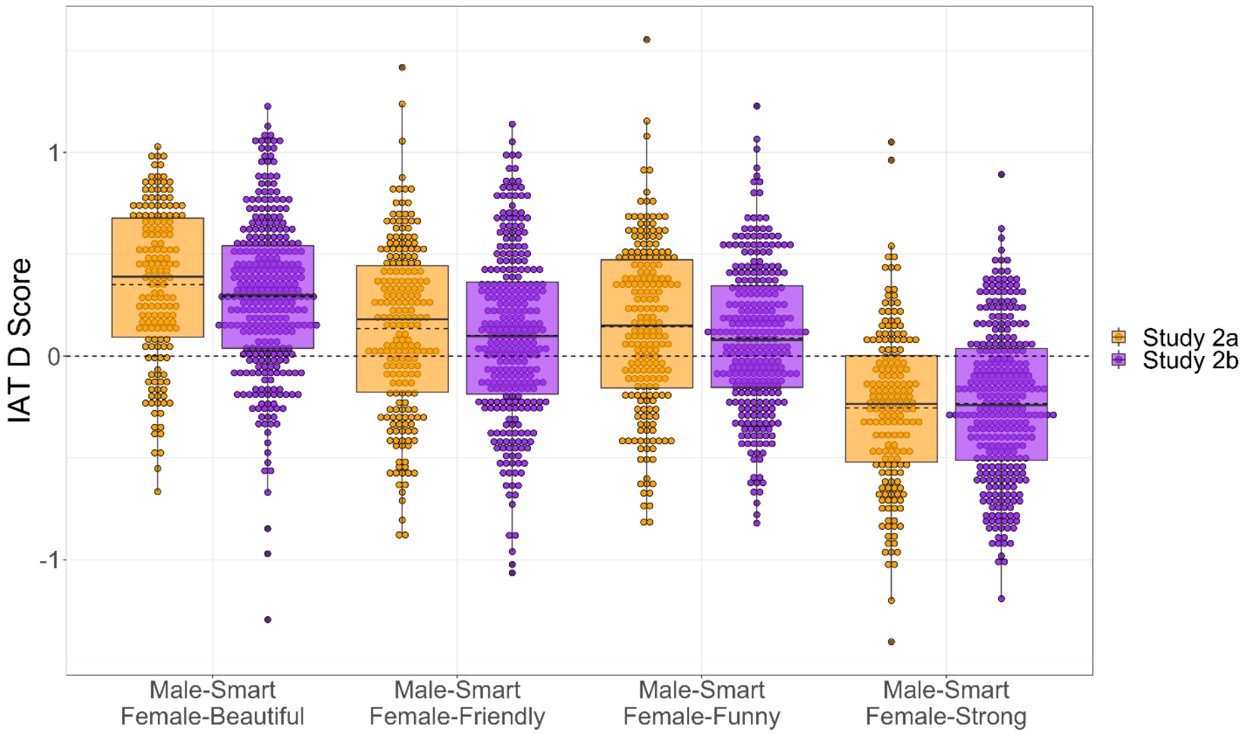

Figure 2. Dot plots (with box plot overlays) of participants' IAT *D* scores in Study 2a and Study 2b. Each dot represents a single participant's *D* score. Solid line = median; dashed line = mean.

Results and Discussion

Implicit Gender-Brilliance Stereotype. As expected, participants showed an implicit *male-super smart* association when the comparison traits were female-typed: *very beautiful* (D_{2a} = 0.35 [0.29, 0.41], D_{2b} = 0.29 [0.25, 0.33]) and *very friendly* (D_{2a} = 0.13 [0.08, 0.19], D_{2b} = 0.10 [0.05, 0.15]), *t*s > 4.13, *p*s < .001, *d*s > 0.24, >60% of *D* scores above 0 (see Figure 2 and OSF). Participants also showed evidence of a gender-brilliance stereotype with the male-typed

correlation between the attitude and stereotype IATs. Residual variance was calculated as 1 – error variance – variance accounted for by the attitude.

comparison trait *very funny* (D_{2a} = 0.14 [0.08, 0.20], D_{2b} = 0.09 [0.05, 0.13]), *t*s > 4.17, *p*s < .001, *d*s > 0.24, >59% of *D* scores above 0. This indicates that the implicit association between *male* and *super smart* is relatively stronger than the implicit association between *male* and *very funny*, an explicitly male-typed trait.

Nevertheless, a reverse association was found for the comparison trait of *very strong*, such that participants associated *male* with *very strong* and *female* with *super smart* (D_{2a} = −0.26 [−0.31, −0.20], D_{2b} = −0.23 [−0.27, −0.19]), *t*s > 8.76, *p*s < .001, *d*s > 0.62, >73% of *D* scores below 0. Thus, only when the comparison attribute was strongly male-typed (and physical) did participants' scores "flip," finally showing an association between *super smart* and *female*. To clarify, this finding does not challenge the existence of an implicit gender-brilliance stereotype; rather, it helps to more precisely position the relative strength of this stereotype. While the tendency to associate brilliance with men is stronger than the tendency to associate humor, friendliness, and beauty with men (as evidenced by average *D* scores significantly greater than 0), it is weaker than the tendency to associate strength with men.

Explicit Gender-Brilliance Stereotype. In Study 2a, participants explicitly associated *super smart* with women more than men (M = 4.10), $t(721) = 3.56$, $p < .001$, $d = .13$, on a *t* test against 4 (the midpoint). Similar results were observed in Study 2b: Participants self-reported a weak but statistically significant association between *super smart* and women (M = 4.05), $t(1137) = 2.15$, $p = .032$, $d = .07$. Thus, similar to Study 1, participants showed no evidence of explicitly endorsing a gender-brilliance stereotype favoring men—in fact, they reported that women are more brilliant than men, which is inconsistent with their implicit stereotypes.

Relation between Explicit and Implicit Gender-Brilliance Stereotypes. Prior to calculating the correlations between the explicit and implicit measures of the gender-brilliance stereotype, we reverse-coded the explicit measure so that it points in the same direction as the IAT, with higher values indicating a gender-brilliance stereotype favoring males. Participants' explicit and implicit stereotypes showed small but significant positive correlations in both Study 2a, $r(720) = .15$, $p < .001$, and Study 2b, $r(1136) = .09$, $p = .004$. The fact that IAT scores correlated positively with self-reports of a gender-brilliance stereotype favoring males provides additional evidence for the validity of our IAT as a measure of the implicit association between *male* and *brilliant* (rather than just *female* and the comparison traits).

Implicit Attitudes Toward Traits (Study 2b). Implicit attitudes (measured with the attitude IATs) favored *super smart* over the comparison traits of *very strong* ($D = 0.36$ [0.31, 0.40]) and *very funny* ($D = 0.15$ [0.10, 0.20]), ts > 5.49, ps < .001. Implicit attitudes toward *super smart* and *very friendly* were matched ($D = 0.01$ [−0.05, 0.06]), $t(292) = 0.29$, $p = .78$, and participants held more positive implicit attitudes toward *very beautiful* than *super smart* ($D = -0.13$ [−0.18, −0.08]), $t(311) = 5.31$, $p < .001$. That only one comparison trait (*very beautiful*) was implicitly perceived as more positive than the trait *super smart* reduces the concern that implicit stereotypes were confounded by valence-matching between gender attitudes and attitudes towards the traits, a conclusion also reinforced by the variance decomposition analyses, which we describe next.

Variance Decomposition Analyses. For all four stereotype IATs, the proportion of independent variance (i.e., the proportion of residual true variance after removing the variance accounted for by the valence of participants' implicit attitudes towards the traits) was significantly different from zero: 0.45 [0.27, 0.60] for the stereotype IAT with *very beautiful* as a comparison trait, 0.40 [0.18, 0.56] for the *very friendly* IAT, 0.41 [0.26, 0.55] for the *very funny* IAT, and 0.26 [0.06, 0.44] for the *very strong* IAT. These results rule out the alternative explanation that the observed implicit gender-brilliance stereotype is simply due to a valence match between people's positive attitudes toward *female* and the comparison traits. More likely, this implicit stereotype reflects meaningful semantic associations.

Conclusions. In summary, the results of Study 2 provide three main conclusions. First, the implicit stereotype that associates being brilliant with men (and other traits with women) is strong—even stronger than the association between men and the male-typed trait *funny*. Second, when asked directly, participants self-report a weak association between *super smart* and women, not men. This result highlights the added value of employing an implicit measure such as the IAT. Third, Study 2b ruled out the possibility that the implicit gender-brilliance stereotype is an artifact of the valence of participants' implicit attitudes toward the categories and traits involved; instead, it indeed appears to be a semantic association.

Having provided evidence for the existence and relative strength of the implicit gender-brilliance stereotype, we expand outwards in Studies 3–5 to test whether this belief generalizes to Black stereotype targets (Study 3), child participants (Study 4), and international participants (Study 5).

Study 3

The photographs used in Study 1 depicted White women and men. In the U.S., stereotypes about White women and men (the high-status majority group) are most similar to stereotypes about women and men *in general* (Ghavami & Peplau, 2013; see also Purdie-Vaughns & Eibach, 2008). In contrast, stereotypes of Black women and men are most different from general gender stereotypes, not only relative to the stereotypes of White women and men but also Latinx, Asian, and other groups. For example, Black women experience less backlash than White women when they behave assertively, in part because assertiveness aligns with Black racial/ethnic stereotypes (Livingston, Rosette, & Washington, 2012). In light of these considerations, participants may not show an implicit gender-brilliance stereotype favoring men when reasoning about Black men and women. Alternatively, it is possible the implicit gender-brilliance stereotype will override such intersectionality effects (i.e., unique effects that arise out of the interaction of various social identities; Crenshaw, 1990), in which case we would expect to see similar IAT scores to what was observed in Studies 1 and 2, even when using Black targets.

To distinguish between these predictions, Study 3 investigated implicit gender-brilliance stereotypes by using pictures of Black females and males (n = 8 each) from the Chicago Face Database (Ma et al., 2015), matched in attractiveness, age, and positive emotionality (see Appendix S1 in the SOM). In this and the following two studies, we use *creative* as a comparison trait because *creative* provides the closest semantic match for *genius* (both reflect intellectual qualities). Additionally, *creative* yielded a result in the middle of the distribution of the comparison traits used in Studies 1 and 2 (i.e., weaker than using *beautiful* but stronger than using *funny*) and may therefore be the best approximation the average strength of the implicit gender-brilliance stereotype favoring men.

Participants

Participants (N = 222; 142 women, 78 men, 2 other) were recruited from UIUC (n = 86, M_{age} = 19.4 years, range = 18–23 years) and NYU (n = 136, M_{age} = 19.3 years, range = 18–23 years). An additional 22 participants (6 from UIUC, 16 from NYU) were excluded from the final sample because of criteria related to IAT scoring (Greenwald et al., 2003) or missed attention checks. The final sample was 40.1% White, 4.5% Black, 32.9% Asian, 9.9% Hispanic, and 12.6% other. Participants received course credit for their participation.

Similar to Study 1 (where we also used *creative* as a comparison trait), the sample size was set a priori to 125 participants in each of the two recruitment venues (2 × 125 = 250), plus an allowance for exclusions. This recruitment plan was followed successfully at NYU. However, the participant pool at UIUC during the semester when this study was conducted was smaller than usual, so we were not able to meet our initial recruitment plan for this venue. Note, however, that even this smaller-than-planned sample is sufficiently powered to detect effects of the same magnitude as those in Study 1. No participants were added after initial data analysis.

Results and Discussion

The mean IAT D score was 0.25 [0.21, 0.29], $t(221) = 11.65$, $p < .001$, $d = 0.78$, corresponding to 78% of participants with D scores above 0. As before, this indicates a strong and widely prevalent implicit stereotype associating *male* with *genius* (and *female* with *creative*). Despite the switch in gender stimuli from White targets to Black targets, participants' scores were consistent in magnitude across Studies 1 and 3, suggesting that the implicit gender-brilliance stereotype generalizes across racial/ethnic boundaries. Additionally, as in previous studies, this stereotype was observed among both women ($D = 0.22$ [0.17, 0.27]) and men ($D = 0.32$ [0.24, 0.39]), ts > 8.41, ps < .001 (see Figure 3), though it was again somewhat stronger in men, $t(218) = 2.19$, $p = .030$, $d = 0.30$. Finally, separate analyses within each sample (UIUC, NYU) also found a significant gender-brilliance stereotype for both women and men participants (see OSF).

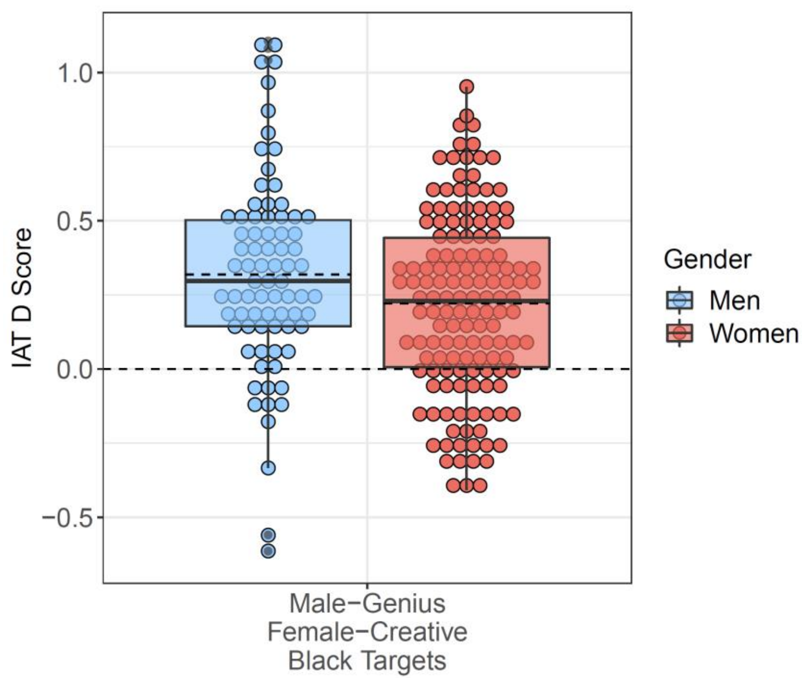

Figure 3. Dot plots (with box plot overlays) of participants' IAT *D* scores in Study 3. Each dot represents a single participant's *D* score. Solid line = median; dashed line = mean.

Study 4

Do *children* also show implicit beliefs associating brilliance with men? This question is important for several reasons. First, implicit beliefs have unique predictive validity in school settings, explaining differences in achievement among children beyond explicit self-reports (Cvencek, Fryberg, Covarrubias, & Meltzoff, 2018; Cvencek, Kapur, & Meltzoff, 2014). Second, certain school subjects and professional fields are generally thought to require more brilliance than others (e.g., mathematics, physics; Leslie, Cimpian, et al., 2015; Storage et al., 2016), with these beliefs even held by elementary school teachers (Heyder, Weidinger, Cimpian, & Steinmayr, 2020). Thus, if children implicitly associate *brilliant* with *male*, this stereotype may be an obstacle to girls' success in fields where they perceive this trait to be valued (e.g., girls may opt out, or be pushed out, of pursuing brilliance-oriented professions). Although previous evidence suggests early acquisition of an *explicit* gender-brilliance stereotype (Bian, Leslie, & Cimpian, 2017), the presence of *implicit* beliefs on this topic has never been investigated in children before. Here, we provide the first test of whether children from two geographically and culturally distinct regions of the U.S. (Illinois and New York) show evidence of an implicit gender-brilliance stereotype.

Participants

Children (N = 103; 51 girls, 52 boys; M_{age} = 9.98 years, range = 9–10 years) were recruited from Urbana-Champaign, Illinois (n = 53) and New York City, New York (n = 50). An additional 3 children (1 from Illinois, 2 from New York) were excluded from the final sample because of criteria related to IAT scoring (Greenwald et al., 2003), a server error, or being outside the targeted age range.

We focused on 9- and 10-year-olds because piloting indicated that 9-year-olds were the youngest group that could reliably complete the same IAT used with adults in Study 1. Although versions of the IAT have been used with younger children (e.g., Cvencek et al., 2011), modifying the IAT for a younger sample would have prevented a direct comparison with adults, a primary interest in this study. Focusing on elementary-school children’s implicit stereotypes, as we do here, is also theoretically motivated, as prior evidence has documented that (other) implicit stereotypes are present even at this early age and predict students’ achievement (Cvencek et al., 2011, 2014).

The sample was 62.2% White, 10.2% Asian, 6.1% Hispanic, 5.1% Black, and 16.3% other. All children were tested in a quiet laboratory environment and received a brief training before completing the IAT (e.g., they were asked to pronounce and define each of the six words used [see Table 1] and were corrected, if necessary).

The sample size was set a priori to 50 usable participants in each of the two recruitment venues (2 × 50 = 100). The samples in the present study are smaller than the samples of adults in the preceding studies because children are more difficult to recruit. Nevertheless, these samples are sufficiently powered to detect effects of the same magnitude as those in Study 1 (which used the same comparison trait—namely, *creative*). No children were added to the sample after initial data analysis.

Results and Discussion

Similar to adults, children showed evidence of a moderate-to-strong implicit stereotype associating *male* with *genius* (and *female* with *creative*), D = 0.24 [0.18, 0.31], $t(102) = 7.43$, $p < .001$, $d = 0.73$, corresponding to 75% of participants' D scores being above 0. This stereotype was present for both girls (D = 0.25 [0.16, 0.33]) and boys (D = 0.24 [0.14, 0.35]), ts > 4.75, ps < .001 (see Figure 4), and did not differ by gender, $t(101) = 0.04$, $p = .97$, $d = 0.01$. Finally, as in the previous studies, the implicit gender-brilliance stereotype was present when examining girls' and boys' responses separately within each sample (Illinois and New York; see OSF).

In summary, the present results reveal that the implicit stereotype associating *genius* with *male* more than *female* (and *creative* with *female* more than *male*) is in place early in life, among elementary-school children growing up in two geographically and culturally distinct regions of the U.S.

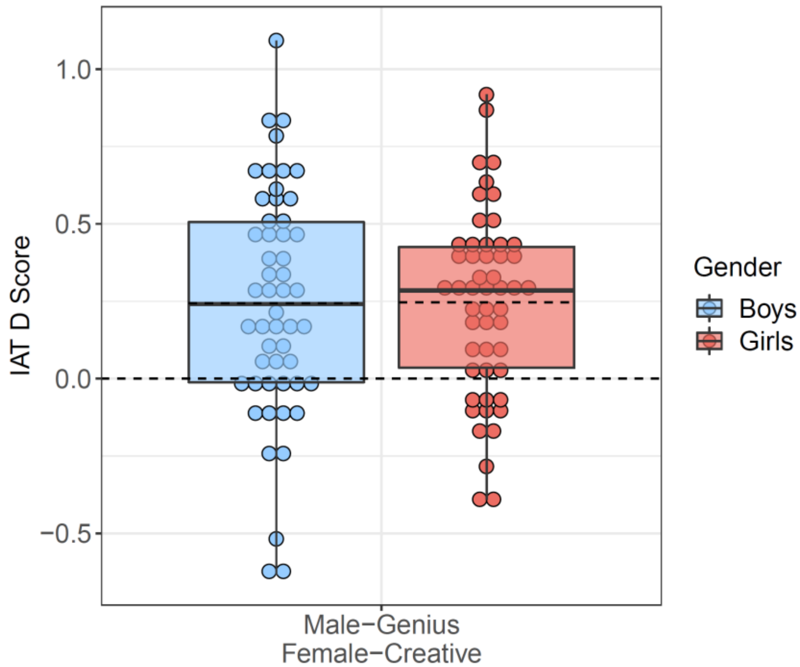

Figure 4. Dot plots (with box plot overlays) of children's IAT *D* scores in Study 4, by gender. Each dot represents a single child's *D* score. Solid line = median; dashed line = mean.

Study 5

Both implicit and explicit gender stereotypes have been shown to vary across cultures (Glick et al., 2000; Miller, Eagly & Linn, 2015; Nosek et al., 2009). As such, Study 5 was intended as a first step toward testing whether the implicit gender-brilliance stereotype extends beyond the cultural context of the Unites States. To ensure as broad a test as possible, we administered our IAT to online samples of participants spanning seven different regions of the world (e.g., Eastern Europe, Latin America and the Caribbean) and tested whether the scores from each of these regions revealed an association between *genius* and *male* (and *creative* and *female*).

Participants

Residents of 78 countries (N = 514; 360 men, 151 women, 3 other; M_{age} = 30.9 years, range = 18–64 years) were recruited through Amazon's Mechanical Turk (see Table S1 in the

SOM for the full list of countries).[5] Before beginning the study, participants completed an English proficiency check, consisting of four questions with answers that would be obvious to anyone who understands English (e.g., "If someone robbed you, how would you feel?", with the answer options being "good," "happy," "energetic," and "sad"). Participants were excluded if they gave the wrong answer to more than two of these four questions; five participants were excluded on this basis. An additional 21 participants were excluded because of criteria related to IAT scoring (Greenwald et al., 2003) or because they had a U.S. IP address. Participants received $1.75 for participating. A majority of the sample was college-educated, with 13.5% reporting having completed "some college," 44.6% reporting having a college degree, and 28.2% reporting having a master's degree. The median household income bracket of the sample was 20,000–29,999 U.S. dollars.

The sample size was set a priori to 500 participants, with an allowance for exclusions. We could not anticipate how various regions would be represented in the sample, so we reasoned that 500 participants would provide a balance between statistical power across regions and cost of recruitment. No participants were added after initial data analysis.

Results and Discussion

The average IAT D score across the entire sample was 0.24 [0.21, 0.27], $t(513) = 15.70$, $p < .001$, $d = 0.69$, corresponding to 77% of participants with D scores above 0. The implicit stereotype associating *genius* with *male* (and *creative* with *female*) thus appears consistently even in participants from cultures outside of the U.S. Moreover, the magnitude of this stereotype was similar to that shown by U.S. participants in the preceding studies. Both women ($D = 0.18$ [0.12, 0.24]) and men ($D = 0.26$ [0.23, 0.29]) showed the implicit *male-genius*/*female-creative*

[5] Mechanical Turk does not support accounts for U.S. citizens residing abroad. Tax information is required by Amazon when registering for an account on Mechanical Turk to verify non-U.S. status.

association, *t*s > 5.97, *p*s < .001, and, as was the case among previous samples of U.S. adults, men's stereotypes were stronger, $t(509) = 2.31$, $p = .021$, $d = 0.20$.

The implicit gender-brilliance stereotype was observed in all seven of the geographic regions sampled in this study (see Figure 5): Eastern and Southeastern Asia ($n = 20$, $D = 0.31$ [0.17, 0.44]), Eastern Europe ($n = 40$, $D = 0.31$ [0.20, 0.43]), Latin America and the Caribbean ($n = 88$, $D = 0.22$ [0.16, 0.28]), Northern and Sub-Saharan Africa ($n = 14$, $D = 0.26$ [0.01, 0.51]), Southern Asia ($n = 160$, $D = 0.16$ [0.10, 0.22]), Western Asia ($n = 20$, $D = 0.26$ [0.07, 0.46]), and Western Europe ($n = 151$, $D = 0.30$ [0.25, 0.35]), *t*s > 2.21, *p*s < .046 (see Table S1 in the SOM for a list of countries in each region).[6] Of note, our relatively small MTurk samples are not necessarily representative of their respective regions and, as such, this study represents a first step rather than the final word on the cross-cultural generalizability of the implicit gender-brilliance stereotypes. Nevertheless, this evidence points to the possibility that the stereotype associating *male* with the trait *genius* (and *female* with the trait *creative*) is a global phenomenon. Future studies should recruit larger or more representative samples from a wider range of countries to assess the cross-cultural generalizability of these findings.

To provide an additional test of the cross-cultural consistency of this stereotype, we ran an intercept-only mixed-effects model in which individuals' IAT scores were nested within region (i.e., the model included a random intercept for region). Consistent with the possibility of cross-cultural consistency, this model revealed that only 1.9% of the variability in IAT scores was due to the region in which a participant resided (95% *CI* = [0.3% to 9.6%]; see OSF).

Finally, in a separate set of exploratory analyses, we examined the country-level

[6] Although Canada did not fit into the seven regions (as defined by the United Nations Statistics Division, 2010), the sample contained 15 Canadian participants as well. As with every other geographic region, Canadian respondents showed an implicit gender-brilliance stereotype, $D = 0.22$ [0.05, 0.39], $t(14) = 2.80$, $p = .014$.

relationship between gender-brilliance stereotyping and the degree of structural inequality between women and men (in terms of economic participation, health outcomes, etc.). Although the present sample of participants was smaller than is typical for such cross-country analyses, and the results should therefore be interpreted with caution, we found that participants from more gender-equal countries exhibited *stronger implicit* but *weaker explicit* gender-brilliance stereotypes (for details, see Appendix S4 of the SOM). If replicated in subsequent work, this pattern of relationships may also provide insight into recent findings of a “gender-equality paradox,” whereby more gender-equal countries also have fewer women pursuing careers in science and technology (Stoet & Geary, 2018; see also Charles & Bradley, 2009; Richardson et al., 2020). Given that high-level intellectual ability is seen as key to success in these careers (Leslie, Cimpian, et al., 2015), the present data suggest that women pursuing science and technology in more gender-equal countries may face more, rather than less, implicit bias—even though an examination of explicit measures of gender bias might suggest otherwise.

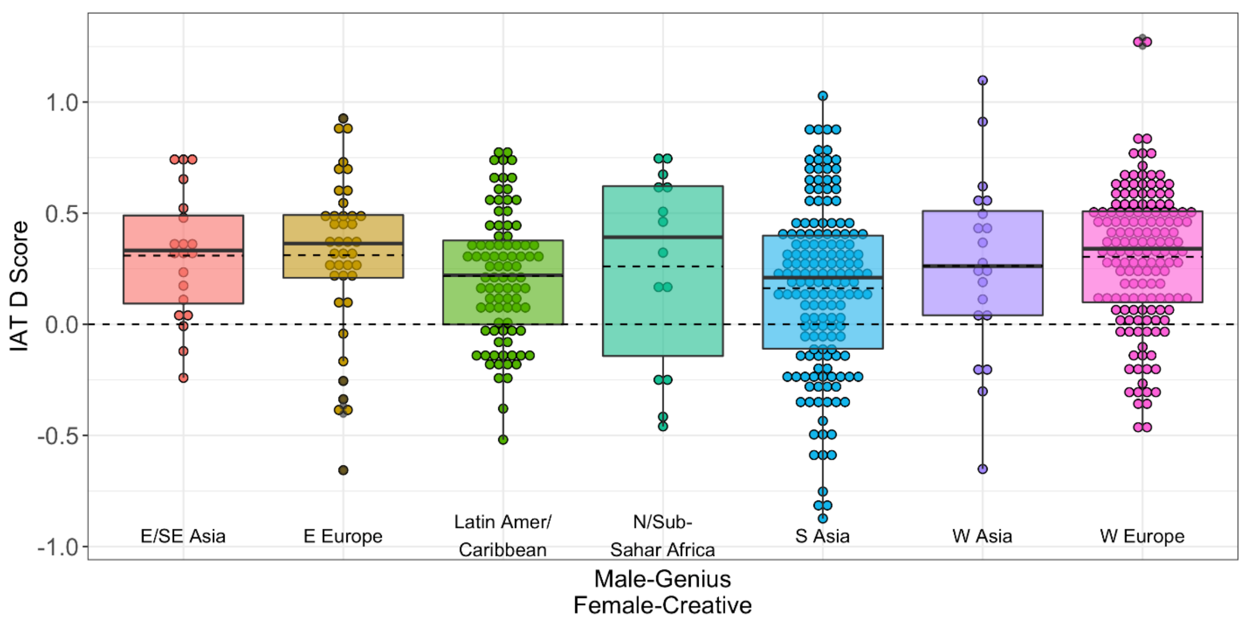

Figure 5. Dot plots (with box plot overlays) of participants’ IAT *D* scores in Study 5, by region. Each dot represents a single participant’s *D* score. Solid line = median; dashed line = mean. E/SE Asia = Eastern and Southeastern Asia; E Europe = Eastern Europe; Latin Amer/Caribbean =

Latin America and the Caribbean; N/Sub-Sahar Africa = Northern and Sub-Saharan Africa; S Asia = Southern Asia; W Asia = Western Asia; W Europe = Western Europe. The regions were defined using a classification scheme from the United Nations Statistics Division (2010).

General Discussion

The present research contributes to a growing literature examining the stereotype that portrays high-level intellectual ability (e.g., brilliance, genius) as a predominantly male quality—a stereotype that holds women back in careers where this ability is valued, both in STEM and beyond (Cimpian & Leslie, 2017). Using implicit measures (namely, the IAT), which were absent from prior investigations of this stereotype, we found a pervasive implicit stereotype associating the traits of brilliance and genius with men more consistently than with women (and other traits, such as creativity or humor, with women more than men). This implicit stereotype was surprisingly robust: It was observed in women and men from the U.S. (Studies 1–3), in women and men from other regions of the world (Study 5), and in American girls and boys as young as 9 years of age (Study 4). Similarly, it was observed when assessed with female and male targets from different racial/ethnic groups (Studies 1 and 3), as well as with generic gender labels instead of images (Study 2), and when assessed in the context of five unique comparison traits, including ones that are explicitly rated as masculine (i.e., funny).

The present studies also ruled out two alternative explanations for the observed implicit stereotype. First, it is unlikely that the gender-brilliance stereotype IAT simply captured positive stereotypes about women: IAT scores correlated positively with explicit measures of the gender-brilliance stereotype (i.e., self-reports of the belief that men are more brilliant than women), as well as with broader measures of explicit bias against women and political conservatism (Studies 1 and 2). Second, it is unlikely that the gender-brilliance stereotype is simply to due to a superficial valence match between participants’ positive attitudes toward women and the comparison traits. Contrary to this alternative, variability in gender-brilliance stereotype IAT scores was not accounted for by the valence of participants' implicit attitudes toward the comparison traits (Study 2b).

Our results using explicit stereotype measures (Studies 1 and 2) further highlight the value of implicit measures to a complete understanding of the gender-brilliance stereotype. When asked *directly* whether they believe that men are more brilliant than women, participants generally reported disagreeing with this idea (Study 1). In fact, participants reported that they associated the quality of being *super smart* with women more than men (Study 2). Interpreting these responses is not entirely straightforward. One possibility is to take them at face value and conclude that there is no explicit gender-brilliance stereotype favoring men, and perhaps also that current explicit stereotypes on this topic favor women. However, this conclusion is at odds with other recent studies that have consistently identified a gender-brilliance stereotype favoring *men* in participants' explicit—albeit more indirect—judgments (e.g., Bian, Leslie, & Cimpian, 2018; Storage et al., 2016). In addition, participants in Study 1 reported that *society* generally believes men to be more brilliant than women, even if they themselves do not.

Thus, an alternative possibility is that participants are reluctant to report endorsing a belief that is socially undesirable (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009). From this perspective, the explicit belief that men are more brilliant than women may be widely endorsed but may surface only in contexts that do not overtly highlight the intergroup nature of evaluations and the possibility of appearing biased (e.g., Storage et al., 2016). The robust evidence of an *implicit* stereotype linking *brilliant* with *male* (see Studies 1–5) is consistent with this argument that explicit gender-brilliance stereotypes favoring men are widely held but socially sensitive. However, the evidence ultimately cannot rule out the above-mentioned possibility that explicit gender-brilliant stereotypes really are weak or reversed relative to implicit ones. More research is needed to understand the wide variability in responses elicited with explicit measures of gender-brilliance stereotypes.

At a broader level, the present findings contribute to theories seeking to explain women's underrepresentation in STEM. Given that success in (some of) these careers is generally assumed to require brilliance (e.g., Leslie, Cimpian, et al., 2015), a widespread implicit stereotype that associates brilliance with men may make it more difficult for women to pursue these fields, whether by leading women to opt out due to lack of belonging or by biasing evaluations of men and women's potential to succeed (Bian et al., 2018). The surprisingly early acquisition of these stereotypes (see also Bian et al., 2017) is an important factor as well: The earlier children start associating brilliance with males, the earlier girls' aspirations may veer away (or be pushed away) from careers that they perceive to rely on this trait.

The Sources of the Implicit Gender-Brilliance Stereotype Favoring Men

Why do people have an implicit gender-brilliance stereotype favoring men? In principle, one possibility is that this stereotype tracks actual gender differences in high-level intellectual ability. This possibility is unlikely: While women and men do differ on average with respect to certain abilities—with some differences favoring men (e.g., spatial reasoning; Lauer, Yhang, & Lourenco, 2019) and others favoring women (e.g., episodic memory; Asperholm, Högman, Rafi, & Herlitz, 2019)—they do not differ in fluid intelligence (e.g., Flynn, 2012). Additionally, although some have argued that men are more variable and thus overrepresented in the extremes of the intelligence distribution (including in the right tail, where "brilliant" minds are found; e.g., Hedges & Nowell, 1995), mounting evidence actually reveals that the extent to which men are more variable than women in intellectual ability varies across time and cultures (e.g., Feingold, 1994; Irwing & Lynn, 2005; for a review, see Charlesworth & Banaji, 2019b). As such, gender-brilliance stereotypes are unlikely to be learned from observing actual differences in brilliance between men and women.

A more likely source of the gender-brilliance stereotype, suggested by the social role theory of stereotype content (e.g., Eagly & Steffen, 1984; Koenig & Eagly, 2014), is the inferences drawn from observing the current unequal distribution of women and men across careers: When people observe unequal gender distributions in fields that emphasize brilliance, they may (incorrectly) infer that these distributions reflect the inherent qualities of men and women (see also Cimpian & Salomon, 2014a, 2014b). More men than women occupy prominent positions in fields that are perceived to require brilliance—such mathematics, physics, and philosophy—both currently (e.g., faculty at top institutions) and historically (e.g., Newton, Einstein, Plato, Aristotle). When exposed to these gender-imbalanced distributions, people may infer that men are simply better suited for careers that require intellectual "firepower" (Hussak & Cimpian, 2015, 2018a, 2018b). According to this account, gender-brilliance stereotypes favoring men develop not from actual gender differences in intellectual ability but rather as an artifact of the structural factors that have historically constrained women's intellectual pursuits.

This mechanism may also explain the consistent levels of the implicit gender-brilliance stereotype across regions of the world (Study 5): While there is, of course, some cross-country variability in the gender composition of various careers, many brilliance-oriented fields remain male-dominated globally (e.g., Charles & Bradley, 2009; Miller et al., 2015). In addition, schoolchildren all over the world are likely to learn about the contributions made by male historical figures such as Newton or Aristotle, which reinforces the association of *brilliant* with *male*. Study 5 is, however, only a preliminary step toward a cross-cultural examination of the gender-brilliance stereotype. Perhaps future investigations with larger and more representative samples will identify greater cross-country and -region variability in this stereotype than we were able to detect with a small and relatively homogeneous sample of Mechanical Turk workers.

From a developmental standpoint, another source of the gender-brilliance stereotype is children's exposure to socialization agents, such as their parents and teachers, who themselves associate brilliance with men and may express this stereotype through their behaviors around children (e.g., Gunderson, Ramirez, Levine, & Beilock, 2012; Musto, 2019). For instance, teachers tend to attribute girls' classroom performance to their greater conscientiousness and boys' to their greater intellectual ability, which does not always "shine through" when boys are inattentive or unmotivated (e.g., J. Cimpian, Lubienski, Timmer, Makowski, & Miller, 2016; Tiedemann, 2000). These beliefs are manifested in the classroom through subtle behaviors (e.g., allowing boys to interrupt girls and monopolize the conversation) that are nevertheless sufficient to reveal to students what their teachers believe about boys' and girls' abilities (Musto, 2019). To the extent that each new generation of children is socialized by adults who exhibit these beliefs and behaviors, it seems likely that gender-brilliance stereotypes will continue to be widespread.

Limitations and Future Directions

Related to some of the discussion above, our studies did not investigate *why* implicit versus explicit measures lead to different conclusions about the presence (and direction) of a gender-brilliance stereotype. Was the implicit-explicit dissociation observed because the link between intelligence and gender is a sensitive topic and people are unwilling to publicly report their stereotyped beliefs on this topic, or because explicit gender-brilliance stereotypes truly differ in content from explicit ones? More generally, questions about how best to account for dissociations between implicit and explicit measures continue to be conceptually and methodologically debated (e.g., Greenwald & Nosek, 2008; Nosek, 2007) and remain unresolved in the literature, not only with respect to gender-brilliance stereotypes. Finding ways to resolve these differences and fully understand how gender-brilliance stereotypes operate at the implicit and explicit level is an important direction for future research.

In future work, we also hope to extend this work by investigating implicit stereotypes about less extreme levels of intellectual ability. For instance, do people implicitly associate traits such as *smart* and *intelligent* with men more than women as well, even though they no longer seem to do so explicitly (Eagly et al., 2019)? Another useful extension of this work would be to replicate our studies with more diverse samples of participants, and to investigate the cross-country generalizability of the gender-brilliance stereotype with larger, more representative samples. Presenting the task to participants in their native language (rather than in English, which for most of them is a second language) may also reveal greater cross-cultural variability by priming participants' unique cultural and linguistic experiences (e.g., Ogunnaike, Dunham, & Banaji, 2010).

Finally, it will be important to gain more insight into the sources of the substantial individual differences that we observed across studies in participants' implicit gender-brilliance stereotyping (see the distribution of the individual data points in Figures 1–5). The measures of sexism and political conservatism included in Study 1 begin to provide an answer to this question, in that they correlated with participants' implicit stereotypes, but collectively they only accounted for 1.6% of the variance in IAT scores. Thus, there remains much to discover about why some people associate *brilliant* with *male*, while others do not.

Implications for Intervention

How might we intervene in light of this research to increase the representation of women in fields that value brilliance? Although gender stereotypes, including implicit gender stereotypes, have the potential for change over the long term (Charlesworth & Banaji, 2019a; 2019b; Eagly et al., 2019), these changes are often slow, especially at the societal level. Thus, more immediate intervention efforts might productively focus on educating members of brilliance-oriented careers about the causes and consequences of implicit and explicit bias (e.g., Carnes et al., 2015; Devine et al., 2017). Educational interventions may also challenge the notion of fixed brilliance being a requirement for success (e.g., Canning, Muenks, Green, & Murphy, 2019; Heyder et al., 2020). A complementary strategy involves changing the way we socialize boys and girls, including modeling beliefs and behaviors that reflect gender neutrality and providing opportunities for exposure to counter-stereotypical role models (e.g., Cheryan, Siy, Vichayapai, Drury, & Kim, 2011; Dennehy & Dasgupta, 2017; Stout, Dasgupta, Hunsinger, & McManus, 2011). Such interventions may be critical in advancing opportunities and access for women in the near term. Over time, their effects may also accumulate and push society's implicit gender-brilliance stereotypes toward neutrality.

Conclusion

To conclude, the present studies show that people *implicitly* conceive of brilliance and genius as male more than female traits, whereas other traits (e.g., creativity, humor, beauty) are conceived of as relatively more female than male. This work suggests new ways of understanding phenomena of theoretical and societal importance (such as gender gaps in STEM and beyond), paving the way for future investigations into the origins of gender-brilliance stereotypes and their consequences for women's and men's career trajectories all over the world.

References

Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Canada: University of Manitoba Press.

Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale, NJ: Erlbaum.

Asperholm, M., Högman, N., Rafi, J., & Herlitz, A. (2019). What did you do yesterday? A meta-analysis of sex differences in episodic memory. *Psychological Bulletin, 145*(8), 785–821.

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes and Tetlock. *Psychological Inquiry*, *15*(4), 279–289.

Batz, C., & Tay, L. (2018). Gender differences in subjective well-being. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being*. Salt Lake City, UT: DEF Publishers.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162.

Bennett, M. (1996). Men's and women's self-estimates of intelligence. *The Journal of Social Psychology*, *136*(3), 411–412.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391.

Bian, L., Leslie, S. J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist*, *73*(9), 1139–1153.

Bian, L., Leslie, S. J., Murphy, M. C., & Cimpian, A. (2018). Messages about brilliance undermine women's interest in educational and professional opportunities. *Journal of Experimental Social Psychology*, *76*, 404–420.

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, *5*(2), eaau4734.

Carnes, M., Devine, P. G., Baier Manwell, L., Byars-Winston, A., Fine, E., Ford, C. E., … Sheridan, J. (2015). The effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine*, *90*(2), 221–230.

Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, *114*(4), 924–976.

Charlesworth, T. E. S., & Banaji, M. R. (2019a). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192.

Charlesworth, T. E. S., & Banaji, M. R. (2019b). Gender in science, technology, engineering, and mathematics: Issues, causes, solutions. *The Journal of Neuroscience*, *39*(37), 7228 – 7243.

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.

Christopher, A. N., & Mull, M. S. (2006). Conservative ideology and ambivalent sexism. *Psychology of Women Quarterly*, *30*(2), 223–230.

Cimpian, A., & Leslie, S. J. (2015). Response to comment on "Expectations of brilliance underlie gender distributions across academic disciplines." *Science, 349*(6246), 391.

Cimpian, A., & Leslie, S. J. (2017). The brilliance trap: How a misplaced emphasis on genius subtly discourages women and African-Americans from certain academic fields. *Scientific American*, *317*, 60–65.

Cimpian, A., & Salomon, E. (2014a). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*(5), 461–480.

Cimpian, A., & Salomon, E. (2014b). Refining and expanding the proposal of an inherence heuristic in human understanding. *Behavioral and Brain Sciences*, *37*(5), 506–527.

Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open*, *2*(4), 2332858416673617.

Crenshaw, K. (1990). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*, 1241–1299.

Cvencek, D., Fryberg, S. A., Covarrubias, R., & Meltzoff, A. N. (2018). Self- concepts, self-esteem, and academic achievement of minority and majority North American elementary school children. *Child Development*, *89*(4), 1099–1109.

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The preschool implicit association test. *Journal of Experimental Child Psychology*, *109*(2), 187–200.

Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math–gender stereotypes in Singaporean children. *Journal of Experimental Child Psychology*, *117*, 73–91.

Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National*

Academy of Sciences, *114*(23), 5964–5969.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.

Devine, P. G., Forscher, P. S., Cox, W. T., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *Journal of Experimental Social Psychology*, *73*, 211–215.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science, 19*(5), 781–789.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2019). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American Psychologist*. http://dx.doi.org/10.1037/amp0000494

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.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles, 30*(1-2), 81–92.

Fiske, S. T., Cuddy, A. 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.

Flynn, J. R. (2012). *Are we getting smarter?: Rising IQ in the twenty-first century*. New York: Cambridge University Press.

Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of Genetic Psychology*, *163*(1), 24–39.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644.

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science*, 174569161982601.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes: Testing three hypotheses. *Psychology of Women Quarterly*, *37*(1), 113–127.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., ... & Annetje, B. (2000). Beyond prejudice as simple antipathy: hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, *79*(5), 763–775.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean? In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 65–82). Hillsdale, NJ: Lawrence Erlbaum Associates.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*(3-4), 153–166.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*(5220), 41–45.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles*, *7*(11), 1097–1108.

Heyder, A., Weidinger, A. F., Cimpian, A., & Steinmayr, R. (2020). Teachers' belief that math requires innate ability predicts lower intrinsic motivation among low-achieving students. *Learning and Instruction*, *65*, 101220.

Hussak, L. J., & Cimpian, A. (2015). An early-emerging explanatory heuristic promotes support for the status quo. *Journal of Personality and Social Psychology, 109*(5), 739–752.

Hussak, L. J., & Cimpian, A. (2018a). Investigating the origins of political views: Biases in explanation predict conservative attitudes in children and adults. *Developmental Science, 21*, e12567.

Hussak, L. J., & Cimpian, A. (2018b). Memory accessibility shapes explanation: Testing key claims of the inherence heuristic account. *Memory & Cognition, 46*(1), 68–88.

Irwing, P., & Lynn, R. (2005). Sex differences in means and variability on the progressive matrices in university students: A meta- analysis. *British Journal of Psychology*, *96*(4), 505–524.

Jajodia, A., & Earleywine, M. (2003). Measuring alcohol expectancies with the implicit association test. *Psychology of Addictive Behaviors*, *17*(2), 126–133.

Jaxon, J., Lei, R. F., Shachnai, R., Chestnut, E. K., & Cimpian, A. (2019). The acquisition of gender stereotypes about intellectual ability: Intersections with race. *Journal of Social Issues*, *75*(4), 1192–2015.

Jost, J. T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science*, *28*(1), 10–19.

Kirkcaldy, B., Noack, P., Furnham, A., & Siefen, G. (2007). Parental estimates of their own and their children's intelligence. *European Psychologist*, *12*(3), 173–180.

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups’ roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*, 371–392.

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *Journal of Experimental Psychology: General*, *146*(2), 194–213.

Kurdi, B., Mann, T. C., Charlesworth, T. E., & Banaji, M. R. (2019). The relationship between implicit intergroup attitudes and beliefs. *Proceedings of the National Academy of Sciences*, *116*(13), 5862–5871.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist, 74*(5), 569–586.

Lauer, J. E., Yhang, E., & Lourenco, S. F. (2019). The development of gender differences in spatial reasoning: A meta-analytic review. *Psychological Bulletin, 145*(6), 537–565.

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*(6219), 262–265.

Livingston, R. W., Rosette, A. S., & Washington, E. F. (2012). Can an agentic Black woman get ahead? The impact of race and interpersonal dominance on perceptions of female leaders. *Psychological Science, 23*(4), 354–358.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Meyer, M., Cimpian, A., & Leslie, S. J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology*, *6*, 235.

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*(3), 631–644.

Morgan, B. L. (1996). Putting the feminism into feminism scales: Introduction of a Liberal Feminist Attitude and Ideology Scale (LFAIS). *Sex Roles, 34*(5–6), 359–390.

Musto, M. (2019). Brilliant or bad: The gendered social construction of exceptionalism in early adolescence. *American Sociological Review, 84*(3), 369–393.

National Science Foundation (2018). *Survey of Earned Doctorates*. Retrieved from https://www.nsf.gov/statistics/srvydoctorates/.

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science, 16*(2), 65–69.

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.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, *6*(1), 101–115.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*(2), 166–180. https://doi.org/10.1177/0146167204271418

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265–292). New York: Psychology Press.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A.G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., … & 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*, 10593–10597.

Ogunnaike, O., Dunham, Y., & Banaji, M. R. (2010). The language of implicit preferences. *Journal of Experimental Social Psychology*, *46*(6), 999–1003.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248.

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, *59*(5-6), 377–391.

Rattan, A., Savani, K., Naidu, N. V. R., & Dweck, C. S. (2012). Can everyone become highly intelligent? Cultural differences in and societal consequences of beliefs about the universal potential for intelligence. *Journal of Personality and Social Psychology*, *103*(5), 787–803.

Raty, H., & Snellman, L. (1997). Children's images of an intelligent person. *Journal of Social Behavior and Personality, 12*(3), 773–784.

Richardson, S. S., Reiches, M. W., Bruch, J., Boulicault, M., Noll, N. E., & Shattuck-Heidorn, H. (2020). Is there a gender-equality paradox in science, technology, engineering, and math (STEM)? Commentary on the study by Stoet and Geary (2018). *Psychological Science*, *31*(3), 338–341.

Rivera, L. A., & Tilcsik, A. (2019). Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review, 84*(2), 248–274.

Schimmack, U. (2019). The implicit association test: A method in search of a construct. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619863798

Schmader, T., Whitehead, J., & Wysocki, V. H. (2007). A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles, 57*(7-8), 509–514.

Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology*, *22*(1), 13–39.

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, *29*(4), 581–593.

Storage, D. (2017). IATScore: Scoring Algorithm for the Implicit Association Test (IAT). *Comprehensive R Archive Network*. Reference manual available at https://cran.r-project.org/web/packages/IATScore/IATScore.pdf

Storage, D. (2018). IATanalytics: Compute Effect Sizes and Reliability for Implicit Association Test (IAT) Data. *Comprehensive R Archive Network.* Reference manual available at https://cran.r-project.org/web/packages/IATanalytics/IATanalytics.pdf

Storage, D., Horne, Z., Cimpian, A., & Leslie, S. J. (2016). The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and African Americans across fields. *PLOS ONE*, *11*(3), e0150194.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*(2), 255–270.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214.

Tetlock, P. E., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior*, *29*, 3–38.

Tiedemann, J. (2000). Gender-related beliefs of teachers in elementary school mathematics.

Educational Studies in Mathematics, *41*(2), 191–207.

United Nations Statistics Division (2010). *The World's Women 2010: Trends and Statistics.* Available from https://unstats.un.org/unsd/demographic-social/products/worldswomen/ww2010pub.cshtml

Vianello, M., & Bar-Anan, Y. (2020). Can the Implicit Association Test measure automatic judgment? The validation continues. *Psychological Science*. https://doi.org/10.1177/1745691619897960

Williams, J. E., & Best, D. L. (1990). *Cross-cultural research and methodology series, Vol. 6. Measuring sex stereotypes: A multination study (Rev. ed.).* Thousand Oaks, CA, US: Sage Publications.