American Psychological Association

American Psychologist


0003-066X/18/$12.00

2018, Vol. 73, No. 9, 1139–1153
http://dx.doi.org/10.1037/amp0000427

# Evidence of Bias Against Girls and Women in Contexts That Emphasize Intellectual Ability

Lin Bian
Stanford University and Cornell University

Sarah-Jane Leslie
Princeton University

Andrei Cimpian
New York University

Despite the numerous intellectual contributions made by women, we find evidence of bias against them in contexts that emphasize intellectual ability. In the first experiment, 347 participants were asked to refer individuals for a job. Approximately half of the participants were led to believe that the job required high-level intellectual ability; the other half were not. A Bayesian mixed-effects logistic regression revealed that the odds of referring a woman were 38.3% lower when the job description mentioned intellectual ability, consistent with the possibility of gender bias. We also found evidence of gender bias in Experiment 2, which was a preregistered direct replication of Experiment 1 with a larger and more diverse sample (811 participants; 44.6% people of color). Experiment 3 provided a developmental investigation of this bias by testing whether young children favor boys over girls in the context of intellectually challenging activities. Five- to 7-year-olds ($N = 192$) were taught how to play a team game. Half of the children were told that the game was for "really, really smart" children; the other half were not. Children then selected 3 teammates from among 6 unfamiliar children. Children's initial selections were driven by ingroup bias (i.e., girls chose girls and boys chose boys), but children subsequently showed bias against girls, choosing girls as teammates for the "smart" game only 37.6% of the time (vs. 53.4% for the other game). Bias against women and girls in contexts where brilliance is prized emerges early and is a likely obstacle to their success.

*Keywords:* gender bias, gender stereotypes, brilliance stereotypes, children

*Supplemental materials:* http://dx.doi.org/10.1037/amp0000427.supp

By many standards, the intellectual achievements of girls and women in the United States have matched, if not surpassed, those of boys and men. Girls make up over half of the children in gifted and talented programs (National Association for Gifted Children, 2015) and get better grades than boys from kindergarten through twelfth grade (Voyer

*Editor's note.* Andrei Cimpian received a 2018 APA Award for Distinguished Early Career Contributions to Psychology. This article is based on an invited presentation at the 126th Annual Convention of the American Psychological Association, held August 9–12, 2018, San Francisco, California.

*Authors' note.* Lin Bian, Department of Psychology, Stanford University, and Department of Human Development, Cornell University; Sarah-Jane Leslie, Department of Philosophy, Princeton University; Andrei Cimpian, Department of Psychology, New York University.

This research was supported by a Robert Larsen Grant for Research in Career Development from the University of Illinois (awarded to Lin Bian) and by National Science Foundation grants BCS-1530669 and BCS-1733897 (awarded to Andrei Cimpian and Sarah-Jane Leslie). The authors' contributions are as follows: Lin Bian, Sarah-Jane Leslie, and Andrei Cimpian conceived the research. Lin Bian conducted the research and analyzed the data, under Andrei Cimpian's supervision. Lin Bian and Andrei Cimpian wrote the article, with input from Sarah-Jane Leslie. Andrei Cimpian serves as the senior author for this work. We are grateful to the individuals and families who participated and to the members of the Cognitive Development Lab at the University of Illinois and New York University for research assistance and helpful discussion. We also thank Joe Cimpian, Madeline Heilman, Francesca Manzi, Allison Master, David Miller, and Yana Toneva for comments on earlier drafts of this article. The data reported in these studies and the analytic syntax are available on the Open Science Framework: https://osf.io/wnesy/?view_only=82be362669944547a81e5fc2c98e2222.

Correspondence concerning this article should be addressed to Lin Bian, Department of Human Development, Cornell University, Beebe Hall 102A, Ithaca, NY 14853, or to Andrei Cimpian, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: linbian@stanford.edu or andrei.cimpian@nyu.edu

& Voyer, 2014). Likewise, women graduate from college at higher rates, as well as from master's and doctoral programs (National Center for Education Statistics, 2017). Given these realities, one might expect women and men to be treated as intellectual equals and be given the same opportunities to pursue intellectually challenging work. The three experiments presented here suggest otherwise. We find consistent evidence of bias against women in contexts that emphasize intellectual ability. Moreover, we show that this bias is present even in elementary school children.

On the surface, Americans' attitudes seem to track the tremendous progress that women have made in terms of their educational and professional attainments. For example, in a 2015 nationally representative poll from the Pew Research Center, 86% of participants indicated that the trait "intelligent" was equally likely to describe men and women (Pew Research Center, 2015). Not far below the surface, however, the stereotypes that associate men rather than women with high-level intellectual ability—a potent subset of the gender stereotypes about competence and agency (e.g., Eagly & Steffen, 1984; Fiske, Cuddy, Glick, & Xu, 2002)—seem to persist. For example, college students use descriptors such as "brilliant" and "genius" two to three times more often in their anonymous evaluations of male than female instructors (Storage, Horne, Cimpian, & Leslie, 2016), and parents search Google for information on whether their sons are "gifted" 2.5 times more often than they do for their daughters (Stephens-Davidowitz, 2014). Similarly, academic letters of recommendation for women contain fewer superlative comments such as "brilliant scientist" or "trailblazer" (Dutt, Pfaff, Bernstein, Dillard, & Block, 2016), as do reviews of grant proposals submitted by women (relative to men with similar track records; Magua et al., 2017).

These gendered notions of intellectual ability predict women's career outcomes. Women are underrepresented relative to men in fields where brilliance is valued—not just in science (e.g., physics) but also in certain fields in the arts and humanities (e.g., philosophy), which are generally not male-typed domains (Cimpian & Leslie, 2015, 2017; Leslie, Cimpian, Meyer, & Freeland, 2015; Meyer, Cimpian, & Leslie, 2015; Storage et al., 2016). A field's emphasis on brilliance predicts its female representation above and beyond other factors that might be relevant, such as the field's typical work–life balance (e.g., Ferriman, Lubinski, & Benbow, 2009), its focus on people versus abstract systems (e.g., Billington, Baron-Cohen, & Wheelwright, 2007), or its selectivity (e.g., Hedges & Nowell, 1995). Thus, the seemingly subtle differences in how people think about the intellectual abilities of women and men translate into macrolevel inequities in their professional trajectories, with women being systematically underrepresented in some of the most prestigious jobs in American society (see also Eagly & Karau, 2002).

The stereotypes that associate brilliance with men give rise to these patterns of underrepresentation in part by shaping women's beliefs and expectations. For example, job advertisements that emphasize exceptional intellectual abilities undermine women's interest because they lead women to (a) doubt their own abilities and (b) expect that *others* will doubt their abilities, which in turn affects their sense of belonging (Bian, Leslie, Murphy, & Cimpian, 2018; Emerson & Murphy, 2015). While the self-doubts are misplaced—there is no compelling evidence of differences in the inherent intellectual abilities of women and men, either at the mean or at the extremes (e.g., Feingold, 1994; Guiso, Monte, Sapienza, & Zingales, 2008; Hyde, 2005; Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Spelke, 2005)—the expectation that others will be biased could be warranted. This is the central issue we investigate here: Is there bias against women in contexts that emphasize high-level intellectual ability? In the context of the present research, we use the term "gender bias" to refer to differential treatment of women and men in contexts where women and men do not differ in the *actual* abilities relevant to those contexts (e.g., the abilities needed to perform a job).

The study of gender bias in professional contexts has a long, and not altogether straightforward, history. For instance, several high-profile studies over the last 20 years have found evidence of discrimination against women in academic science (e.g., Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Steinpreis, Anders, & Ritzke, 1999; Wenneras & Wold, 1997), but other analyses have in some cases not supported this conclusion (for a review, see Ceci & Williams, 2011). In fact, a 2015 experiment found a bias *favoring* highly qualified women (vs. equally qualified men) for tenure-track positions in several science fields (Williams & Ceci, 2015). These contradictory findings highlight the context-dependent nature of bias, a conclusion reinforced by a recent meta-analysis of employment decisions (Koch, D'Mello, & Sackett, 2015). Moderation tests in this meta-analysis found that, for example, bias against women was strongest in male-dominated fields (consistent with the work on role congruity and perceived "fit"; Eagly & Karau, 2002; Heilman, 2012) and when the competence of the potential employee was ambiguous (see also Heilman, 2012).

The present studies investigate gender bias in a specific type of context—namely, in circumstances that explicitly emphasize raw intellectual talent. It is important to examine whether gender bias is present in these contexts for both practical and theoretical reasons. Practically, evidence of such bias would reveal an obstacle to women's participation and advancement in many prestigious, high-status careers. The idea that brilliance is essential for success is common not only in many STEM fields (i.e., science, technology, engineering, and mathematics) but also in fields such as philosophy and music composition (Leslie, Cimpian, et al., 2015). Young women hoping to pursue such "genius fields" might encounter bias not just from academics in these fields (e.g., in a university setting) but also from members of the general public (e.g., parents, teachers), who generally share academics' beliefs on this topic (Meyer et al., 2015). Even beyond these fields, brilliance may be seen as a prerequisite for many prestigious positions, top awards, and so on; the more rarefied the professional honor, the stronger the expectation may be that its recipient possesses exceptional intellectual ability, regardless of field. If gender bias were present in these contexts, it would impede women's advancement into the upper echelons of their professions. Finding evidence of gender bias in contexts that value brilliance would be theoretically informative as well, as it would speak to a potential mechanism underlying women's underrepresentation in "genius fields" (Leslie, Cimpian, et al., 2015). Prior work has suggested that the culture of these fields tends to undermine women's interest (Bian et al., 2018), but the extent to which others' biases are part of the explanation for this phenomenon is currently unclear.

In addition to focusing on this type of context, the present investigation of gender bias is innovative in two other respects. First, whereas most studies of gender bias in professional contexts have focused on the hiring decision per se, such bias—if present—may be even stronger at earlier stages of the recruitment process, when the pool of candidates is being assembled. Across academia and industry, one of the most common recruitment practices consists of eliciting *referrals* of suitable candidates (e.g., Gërxhani, & Koster, 2015; Zottoli & Wanous, 2000). The act of making a referral (that is, of identifying a familiar person who matches the requirements of a particular job) relies heavily on memory retrieval, a cognitive process that is susceptible to distortion by stereotypes (e.g., Fyock & Stangor, 1994). As a result, the referrals provided in contexts that emphasize brilliance may skew toward the group favored by the "brilliance = men" stereotype, biasing the recruitment process *before* the hiring decision. We thus investigated whether there is bias against women in referrals for jobs described as requiring high-level intellectual ability (Experiments 1 and 2).

The second innovative aspect of the present studies is that we investigated gender bias developmentally. Although it may seem surprising, the processes that discourage women's participation in certain prestigious careers (such as those that prize brilliance) may have their roots early in life (e.g., Bian, Leslie, & Cimpian, 2017; Cheryan, Master, & Meltzoff, 2015; Eccles, 1994; Gunderson, Ramirez, Levine, & Beilock, 2012). Indeed, the stereotypes that associate natural brilliance with males arise as early as kindergarten and first grade (Bian et al., 2017). We thus tested whether contexts that emphasize intellectual ability elicit gender bias even among young children. Biased comments or behaviors from peers may exclude girls from certain domains of activity, depriving them of the opportunity to develop the skills needed to pursue careers in those domains later in life. Thus, we investigated whether children aged 5 to 7 exhibit bias against girls in the context of unfamiliar activities that are described as requiring high-level intellectual ability; these activities serve as experimental analogs of important real-world domains such as mathematics and natural science (Experiment 3).

## Experiment 1

Experiment 1 provided a first test of the hypothesis of bias against women in contexts that emphasize high-level intellectual ability. Participants were asked to refer acquaintances for a job; half of the participants were told that the job was for brilliant individuals, while the other half were not. If there is gender bias in contexts where intellectual ability is salient, we should see that participants are less likely to refer women when they think the job requires brilliance.

### Method

**Participants.** Participants were recruited via Amazon's Mechanical Turk ($N = 347$; 59.7% female; mean age = 35 years). An additional 15 participants were excluded from the experiment because they resided outside of the United States ($n = 12$),[1] reported during debriefing that they had not paid attention ($n = 2$), or did not provide any referrals ($n = 1$).

**Research ethics.** All procedures for this and subsequent experiments were approved by the institutional review board of a large public university in the United States.

**Manipulation.** Participants were asked to imagine that they are working for "a big company" that is "looking to fill a couple of positions in their workforce." They were then randomly assigned to read one of two job descriptions. Participants in the experimental condition read a brief description that emphasized the candidates' intellectual abilities (e.g., "high IQ," "superior reasoning skills," "natural intelligence"; for the full text, see Table S1 in the online supplemental material). In contrast, participants in the control condition read a job description that emphasized the candidates' motivation, which is arguably a requirement for most jobs (e.g., "highly motivated," "superior commitment," "consistent effort"). This description was adapted from prior work (Bian et al., 2018), where it was found to be gender-neutral (i.e., equally applicable to men and women).

**Memory check.** Immediately after reading the job description, participants' memory of it was tested. Specifi-

[1] We excluded non-U.S. Mechanical Turk participants because of uncertainty about (a) whether they are proficient in English and (b) whether they share the same cultural attitudes as U.S. participants (in this case, gender attitudes).

cally, participants were asked to check which attributes from a list of eight had been mentioned in the job description. Of the eight characteristics, four were relevant to the "brilliance" job and four to the control job. Participants whose answers suggested they had not read the job descriptions carefully enough to answer this basic question were excluded from further analyses ($n = 52$ beyond the final 347). Specifically, participants were excluded if they (a) selected none of the characteristics relevant to their condition, or (b) selected one or more characteristics relevant to the other condition (e.g., "high IQ" in the control condition).

In addition to serving as a memory check, this part of the procedure served to reinforce the manipulation: Participants were then shown the correct answers, which all pertained to intellectual ability in the experimental condition (e.g., "high IQ," "natural intelligence") and motivation in the control condition (e.g., "highly motivated," "superior commitment").

**Referral measure.** Next, participants were asked to recommend someone they know who meets the requirements of the job, regardless of whether they already have a job. To mask the purpose of the study, we required participants to provide several pieces of information about the person they were referring in addition to their gender (e.g., their first name, their age, their relationship to the participant). After completing this information, participants were asked to refer a second person and were reminded of the job description before being allowed to do so. Asking participants to make decisions sequentially (vs. simultaneously), as we did here, may provide a more sensitive measure of bias (see Brooks & Purdie-Vaughns, 2007). The two referrals were included in our analyses as separate data points (see the Results and Discussion section for more detail).

Three participants gave exactly the same responses for their first and second referrals (same name, age, gender, relationship to the participant, etc.). The second referral was excluded from the analyses for these participants, but the first was kept.

**Brilliance stereotype measure.** To explore how participants' referrals relate to their gender stereotypes, at the end of the sessions we administered eight items describing purported differences between the intellectual abilities of women and men. Four of the items pertained specifically to gender differences in overall amount of intellectual ability (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*). Because social desirability concerns might suppress participants' agreement with these items, we also administered four items that spoke more generally (and obliquely) about intellectual differences between men and women (e.g., "Even though it may not be politically correct to say it, males and females might be naturally suited for different kinds of intellectual activities"). A parallel analysis in an exploratory factor analysis of the eight items suggested a one-factor solution (eigenvalue = 4.91; 61.4% of the variance explained). Thus, these items were averaged into a composite stereotype score ($\alpha = .90$).

**Additional measures.** Several additional measures (e.g., participants' justifications for their referrals) were collected and are described in Appendix S1 in the online supplemental material.

**Analytic strategy.** In all three experiments, we submitted participants' referrals (0 = *man*, 1 = *woman*) to multilevel mixed-effects logistic regressions. The two referrals provided by each participant were included as separate data points (Level 1), nested within subject (Level 2); we did not sum or average responses within a participant. We fitted the mixed-effects logistic regressions within a Bayesian framework, using the *brms* package for R (Bürkner, 2016; see also Stan Development Team, 2017b).

Bayesian statistics are becoming increasingly common in psychological science (van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017) because they present multiple advantages over conventional frequentist statistics. These advantages have been laid out comprehensively in many recent articles (e.g., Dienes & Mclatchie, 2018; Kruschke & Liddell, 2018; Wagenmakers et al., 2018); here, we list those benefits that we see as most important. First, Bayesian statistics enable researchers to draw inferences about the parameters of interest in light of the data they have collected, which is the typical aim of scientific research. In contrast, frequentist statistics supply the probability of the *data* given the hypothesis being tested. Thus, the very logic of frequentist statistical inference is somewhat at odds with the goal of scientific practice, which is to assign probabilities to hypotheses, not data sets. Second, Bayesian statistics allow researchers to make easily interpretable, intuitive statements about the probability that an estimated parameter falls in a certain interval given the observed data: The probability that a parameter falls in its Bayesian 95% credible interval is 0.95. Frequentist confidence intervals do not allow such statements (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). Below, we report Bayesian 95% credible intervals for all parameters of interest. Third, and related to the second point, Bayesian credible intervals are useful because they allow inferences about the relative credibility of the values within them: Values close to their center are more credible than values close to their limits. As a result, it is relatively unimportant whether the credible interval for a parameter estimate just crosses versus just avoids crossing 0. In either case, 0 is a low-credibility value of the parameter because it is in the tail of its credible interval. Fourth, because of the properties above, Bayesian (vs. frequentist) statistics are better aligned with the goals of the widely endorsed "new statistics" (e.g., Cumming, 2014), which recommend placing the focus on estimating parameters (and the uncertainty around them) rather than significance testing (Kruschke & Liddell, 2018). For completeness, however, we also computed frequentist versions of the multilevel mixed-effects logistic regressions (see Tables S6–S8, available in the online supplemental material); their results are in agreement with the results of the Bayesian analyses.

**Open science practices.** The data and analytic syntax for all experiments are available on Open Science Framework (OSF; https://osf.io/wnesy/?view_only=82be362669944547a81e5fc2c98e2222).

## Results and Discussion

The primary test of whether participants exhibited gender bias in this experiment was the contrast between the experimental and control conditions: Were participants less likely to refer women when the job was said to require brilliance than when it was not? However, gender bias might also be revealed by comparisons to the gender-neutral 50% threshold: Did participants' referrals depart substantially from a 50/50 gender split? Although informative in some ways, such departures from 50/50 do not necessarily signal prejudiced gender attitudes and are thus more difficult to interpret. Because social and professional networks tend to be segregated by gender (e.g., Kolker, 2017), participants' referrals are likely to consist of a majority of people of their own gender even if they harbor no negative views about the other gender.[2] As a result, our conclusions about gender bias in this research are based primarily on the first test above (i.e., is there a condition difference in referrals?), which avoids the interpretive difficulties associated with the second test (i.e., is there a numerical imbalance in referrals?). Whenever we highlight deviations from a 50/50 gender split as relevant to gender bias, we do so only in the aggregate (i.e., adjusting for participant gender): A numerical imbalance in referrals may be most indicative of prejudice when the homophilic biases in the social/professional networks of men and women are allowed to cancel each other out.

Were participants less likely to refer women when the job required brilliance than when it did not? To answer this question, we performed a Bayesian multilevel mixed-effects logistic regression on the gender of participants' referrals in each referral round (0 = *man*, 1 = *woman*; Level 1), nested within participant (Level 2). The model included condition (0 = *control*, 1 = *experimental*; Level-2 predictor), participants' gender (0 = *man*, 1 = *woman*; Level-2 predictor), their stereotype scores (Level-2 predictor), plus all possible two- and three-way interactions between these predictors as fixed effects and a random intercept for participants. For all fixed effects in the model, we set a weakly informative prior: a normal distribution with a mean of 0 (i.e., no relationship) and a *SD* of 1.[3] For the random effect, the prior was the *brms* default (a Student's *t* distribution with a mean of 0, a *SD* of 10, and 3 degrees of freedom). All means reported below are estimated margins from this mixed-effects model, which adjust for the values of all other predictors.

Consistent with the possibility of gender bias in contexts that emphasize intellectual ability, participants were less likely to refer a woman in the experimental condition (40.5% [35.1 to 46.1%] female referrals) than in the control condition (52.5% [46.8 to 58.0%]), $b = -0.483$ [−0.807, −0.165] (Figure 1). The odds of referring a woman (vs. a man) were 38.3% lower when the job description mentioned brilliance.

It is also noteworthy that the credible interval for the experimental condition (35.1 to 46.1% women referred) was entirely below the neutral 50% threshold. This imbalance speaks to the presence of bias in the second sense defined above, especially when contrasted with the fact that the credible interval for the control condition (46.8 to 58.0% women referred) spanned 50%.

The model also revealed a relationship between the gender of the participants and the gender of the referrals, $b = 0.967$ [0.635, 1.301]. Women were more likely to refer a woman than men were (56.3% [51.2 to 61.2%] and 32.9% [27.3 to 38.8%] female referrals, respectively).[4] However, there was no credible interaction between participant gender and condition, $b = -0.147$ [−0.790, 0.490], suggesting that women and men exhibited similar levels of brilliance gender bias (see also Moss-Racusin et al., 2012).

Participants' stereotypes also negatively predicted the probability of referring a woman, $b = -0.087$ [−0.181, 0.007].[5] With a 1-point increase (on a 1–9 scale) in participants' agreement with the statements claiming differences between the intellectual abilities of men and women, the odds of referring a woman (vs. a man) decreased by 8.4%. This relationship did not vary by participant gender, $b = 0.005$ [−0.183, 0.193], or condition, $b = -0.080$ [−0.261, 0.100]. The fact that higher-

[2] Even so, this type of bias is likely to exacerbate gender gaps in real-world professional contexts where one gender is already underrepresented.

[3] This prior is *informative* only by contrast with a uniform (flat) prior distribution, which is typically labeled "uninformative." We did not use a uniform prior because it can cause estimation issues (Stan Development Team, 2017a). The normal(0,1) prior is *weakly* informative because it still allows a wide range of parameter values, but it makes estimation more tractable by ruling out unreasonable parameter values.

[4] The fact that men's referrals were overall more skewed toward their own gender than women's were reinforces the practical importance of considering the gender of the individuals from whom referrals are solicited.

[5] Although the credible interval of this coefficient crosses 0 by a small margin (as does the analogous coefficient in Experiment 2), values in the tails of Bayesian credible intervals—such as 0 in this case—are less probable than values in their center. This is one of the important ways in which the interpretation of Bayesian credible intervals departs from that of (frequentist) confidence intervals, for which crossing 0 renders a result "not significant."

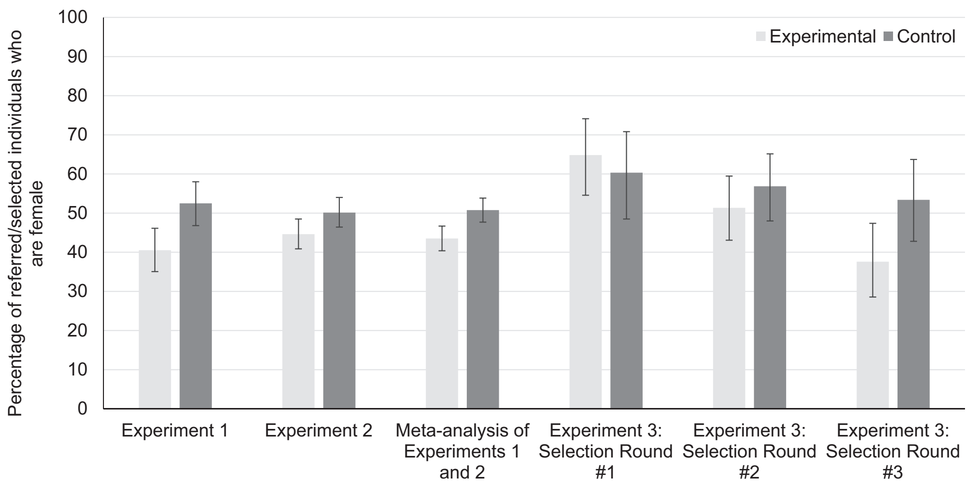


*Figure 1.* The percentage of females referred (Experiments 1 and 2) or selected (Experiment 3) in the experimental and control conditions. The error bars represent 95% credible intervals. The percentage estimates and the credible intervals were calculated as marginal effects in Bayesian mixed-effects multilevel logistic regressions.

stereotyping participants were less likely to refer a woman regardless of the content of the job description suggests that our stereotyping measure may track, at least in part, individual differences in a broader bias against women's competence that applies across contexts. No other coefficients in this model indicated credible relationships (for full model output, see Table S2, available in the online supplemental materials).

## Experiment 2

Experiment 2 was a preregistered direct replication of Experiment 1. Our goal was to assess the replicability and generalizability of the gender bias uncovered in Experiment 1 with a larger, more ethnically diverse sample.[6] We did not have strong a priori predictions about whether the gender bias documented in Experiment 1 varies across ethnic groups; Experiment 2 allowed us to explore whether it does.

### Method

**Preregistration.** The preregistration for this experiment is available on AsPredicted.org: https://aspredicted.org/ax7vk.pdf.

**Participants.** Our sample was recruited via Qualtrics Online Panels and consisted of 811 U.S. participants (55.2% women; mean age = 41 years; 55.4% non-Hispanic White, 31.9% non-Hispanic Black, 4.7% Asian, 4.2% Hispanic, 0.5% American Indian/Alaskan Native).

Three of the exclusion criteria from Experiment 1 were preregistered and were used by Qualtrics Online Panels to automatically terminate the study for any participant who failed them: (a) explicitly reporting inattention during debriefing, (b) not filling out the main dependent variable (that is, the gender of a referral), and (c) not reading the job description carefully (as indicated by failing to select at least one of the characteristics mentioned in the job description, or selecting one or more characteristics relevant to the other condition).

When reading over the responses received from Qualtrics Online Panels after these automatic exclusions, we noticed that (a) some participants filled out gibberish or otherwise made it clear that they did not take the survey seriously (e.g., they referred President Lincoln for the job; $n = 9$), and (b) some participants had identical information for their first and second referrals ($n = 25$). Even though we had not preregistered these criteria, we judged there was sufficient justification for excluding the nine participants who filled in nonsense and for omitting the second referral of the 25 participants who provided duplicate referrals (while retaining their first referral). To clarify, we did not preregister these criteria because these problems had not been salient in the Mechanical Turk sample from Experiment 1. After noticing these issues in the Qualtrics sample and deciding to exclude participants based on them, we went back to the Mechanical Turk sample and applied them there as well to ensure consistency. No participants in the Mechanical Turk sample had filled in nonsense, so no additional exclusions were made on this basis. However, we did discover three duplicate referrals, as described in Experiment 1.

[6] We did not record participant race in Experiment 1. Although Mechanical Turk samples are typically more diverse than undergraduate populations, they are still primarily White (e.g., Levay, Freese, & Druckman, 2016).

We note one other sample-related departure from the preregistration. Our sample was originally intended to match the U.S. Census with respect to the breakdown of racial/ethnic groups, thereby approximating a nationally representative sample. Qualtrics Online Panels was not ultimately able to provide the sample demographics agreed upon. Although more diverse than the sample for Experiment 1, the sample for Experiment 2 oversampled African Americans and undersampled other groups relative to the U.S. Census (particularly Latino/as).

## Results and Discussion

Participants' referrals were analyzed with a Bayesian multilevel mixed-effects model structured as in Experiment 1, except with the addition of an ethnicity variable (0 = *participant of color*, 1 = *White participant*; Level-2 predictor) and all its interactions as fixed effects. Rather than updating the priors based on the results of Experiment 1, we made the conservative choice to retain the same normal(0,1) prior, which assumed no relationship between the predictors and the dependent variable.

Replicating the results of Experiment 1, participants were less likely to refer a woman for a job requiring brilliance (experimental: 44.6% [40.9 to 48.5%] vs. control: 50.1% [46.4 to 54.0%] female referrals), $b = -0.222$ [−0.444, −0.007]. The odds of referring a woman (vs. a man) decreased by 19.9% when the job description mentioned intellectual ability. The credible interval for the experimental condition (40.9 to 48.5% women referred) was again below the neutral 50% threshold, whereas the credible interval for the control condition (46.4 to 54.0% women referred) was symmetrical around 50% (see also Figure 1).

As in Experiment 1, women were more likely to refer a woman than men were (60.7% [57.3 to 64.1%] and 31.9% [28.2 to 35.6%] female referrals, respectively), $b = 1.197$ [0.975, 1.417]. The model also revealed that White participants were less likely to refer a woman (43.9% [40.4 to 47.4%] female referrals) than were participants of color (52.2% [48.1 to 56.3%]), $b = -0.333$ [−0.550, −0.116].[7] Neither of these demographic variables moderated the brilliance gender bias revealed by the difference between the experimental and control conditions (see Table S3, available in the online supplemental material).

As in Experiment 1, high-stereotyping participants were somewhat less likely to refer a woman, $b = -0.051$ [−0.109, 0.008]. Unlike in Experiment 1, however, the relationship between participants' stereotypes and the likelihood of referring a woman differed for male versus female participants, $b = 0.117$ [0.001, 0.233]. Stereotyping was negatively related to the likelihood of referring a woman for male participants, $b = -0.116$ [−0.202, −0.032], but not for female participants, $b = 0.002$ [−0.078, 0.083].[8]

In summary, Experiment 2 replicated the finding of a bias against women in contexts that emphasize intellectual ability in a larger, more diverse sample. However, the magnitude of the bias was smaller in this experiment compared with the first. To more precisely estimate the magnitude of the bias observed in this experimental context, we meta-analyzed the two studies.

## (Mini) Meta-Analysis of Experiments 1 and 2

We meta-analyzed Experiments 1 and 2 with a multilevel model structured as in Experiment 1, with two additional random effects: (a) a random intercept for participants nested within experiment, and (b) a random slope for experimental versus control condition. The meta-analytic estimate of gender bias in referrals for brilliance jobs was −0.292 [−0.470, −0.111]. Participants were less likely to refer a woman when the job description mentioned brilliance (43.5% [40.4 to 46.7%] female referrals) than when it did not (50.8% [47.7 to 53.8%]). The odds of referring a woman (vs. a man) for the brilliance job were 25.3% lower than for the control job.

# Experiment 3

The first two experiments provide support for the hypothesis of a bias against women in contexts where intellectual ability is salient. In Experiment 3, we explored the developmental roots of this bias. If present, bias against girls in settings where success is thought to depend on intellectual talent (e.g., math, science; Chestnut, Lei, Leslie, & Cimpian, 2018) could deprive girls of valuable opportunities to develop their skills and, ultimately, may narrow the range of careers they can consider.

## Method

**Participants.** Children between the ages of 5 and 7—the period during which the "brilliance = men" stereotype first emerges (Bian et al., 2017)—were recruited in a small city in the Midwestern United States ($N = 192$; 50.0% girls; mean age = 6.47 years). Children were recruited from a database of families interested in participating in research studies and from local schools and daycares. Written informed consent was obtained from each child's parents prior to the testing session. Children were tested either in a university lab ($n = 57$) or in a quiet room at their school

[7] Similar to the point we made above, the fact that White participants' referrals were male-biased whereas those of participants of color were more gender-neutral reinforces the practical importance of soliciting referrals from a racially diverse group of individuals to achieve equity in recruitment outcomes.

[8] None of the variables listed in the "secondary analyses" section of the preregistration moderated the experimental effect, so they are not reported here (see Appendix S1 in the online supplemental material for additional detail).

($n$ = 135). Demographic information was available for 71% of the families. The ethnic composition of the sample mirrored that of the community in which this research was conducted: 69% of the children were non-Hispanic White, 9% Asian American, 5% non-Hispanic Black, 3% Latino or Hispanic, 1% other (parents did not specify), and 12% multiracial. The median yearly household income was $90,000 (range = $4,500 to $185,000). Seventy-seven percent of the parents in the sample had at least a bachelor's degree. Children's ethnicity and socioeconomic status did not credibly moderate the results reported below; however, these null effects should be interpreted with caution because our study was not designed, and was likely underpowered, to detect moderation by these demographic variables. Two additional children were tested but excluded from the sample because they refused to complete the study.

**Manipulation.** Children were introduced to an unfamiliar team game. After seeing a picture of the game (selected from the Novel Object and Unusual Name Database; Horst & Hout, 2016) and hearing the rules of the game (see Table S4, available in the online supplemental material), children in the experimental condition were told that the game was "only for children who are really, really smart." They were then shown pictures of six unfamiliar children (three boys and three girls, randomly arranged in front of the participant), all at once, and told that they would have to select teammates. Because of the strong ingroup bias present in this age group (e.g., Bian et al., 2017; Dunham, Baron, & Banaji, 2016; Halim, Ruble, Tamis-LeMonda, Shrout, & Amodio, 2017; Powlishta, 1995; Shutts, Pemberton, & Spelke, 2013; Yee & Brown, 1994; Zosuls et al., 2011), it was important to allow children to select *several* teammates, which should give them the opportunity to act on other biases as well, if present. Thus, we asked children to select three of the six children, one at a time, to be on their team. To ensure that children encoded the manipulation information, the experimenter asked children to recall who was said to be good at the game (i.e., "children who are really, really smart") before they made any teammate selections; children's answers were corrected when appropriate. After the first teammate selection, children were also provided with a reminder of the relevant ability messages (i.e., that the game is "for children who are really, really smart"). After selecting three teammates for the first game, children were then shown another game for "really, really smart children," with different rules, and selected another three teammates from a new set of six unfamiliar children.

Children in the control condition went through exactly the same procedure (choosing three teammates for each of two games), except that they were not told that the games were for children who are really smart and thus did not receive any memory questions or reminders either.

The pictures of potential teammates were found through Google Images and were selected on the basis of a norming study ($n$ = 29 participants on Mechanical Turk) such that the boys and girls were matched in perceived age and attractiveness. In this study, the pictures depicted children who were White because recent work suggests that, at least early in development, the "brilliance = men" association is clearest when judging White targets (Boston, Lei, Chestnut, Vraneski-Shachnai, & Cimpian, 2018; see also General Discussion). The pairing between these pictures and a particular game was counterbalanced across participants.

**Brilliance stereotype measure.** After the main portion of the study, we administered a brief measure of gender stereotypes about intellectual ability adapted from Bian et al. (2017). Children were shown six pairs of (White) individuals, one pair at a time, and had to choose which of the individuals in each pair was "really, really smart." The first two pairs consisted of individuals of the same gender, which served to conceal the purpose of this task as a measure of children's beliefs about gender. The last four pairs consisted of a woman and a man, and the proportion of men selected on these trials served as a measure of children's gender stereotypes. Because of experimenter error, one participant was not administered the stereotype measure; this child's stereotype score was imputed on the basis of his gender and age.

**Debriefing.** At the end of the sessions, children received a thorough debriefing that emphasized the importance of effort and learning.

## Results and Discussion

As in Experiments 1 and 2, the primary test of whether participants exhibited brilliance gender bias was the comparison between the experimental and control conditions. If this bias is present, we would expect to see a main effect of condition (experimental vs. control), with children less likely to select girls as teammates when the game is said to require brilliance than when it is not. However, as discussed previously, children's teammate selections might also be influenced by ingroup bias, which should lead them to choose children of their own gender regardless of condition. If ingroup bias is present, we would expect to see a main effect of gender, with girls choosing more girls than boys do. It is also possible that these two biases will vary in strength across the three selection rounds, in which case we should see statistical interactions between condition and selection round (for the brilliance bias) and gender and selection round (for ingroup bias). For example, if a certain type of bias is present throughout the task but weaker in earlier or later selection rounds, we should observe both a main effect of the relevant factor (condition for brilliance bias, gender for ingroup bias) and an interaction between this factor and selection round. As an alternative, a bias might be present in some selection rounds but absent in others, in which case we should observe an interaction between the relevant factor and selection round but perhaps no main effect of this factor. Our prediction was that both brilliance and ingroup biases would influence children's responses, but we did not have a priori expectations regarding when in the task these biases would be observed.

The gender of the teammate chosen in each selection round (0 = *boy*; 1 = *girl*) was analyzed with a Bayesian multilevel mixed-effects logistic regression (Level 1: children's selections on each round, Level 2: child). Condition (0 = *control*, 1 = *experimental*; Level-2 predictor), children's gender (0 = *boy*, 1 = *girl*; Level-2 predictor), their age (Level-2 predictor), their stereotype score (Level-2 predictor), and all possible interactions were used as predictors. The fixed-effects model also included selection round (first, second, or third; Level-1 predictor), plus interactions, to account for the possibility that the strength of brilliance and ingroup biases might vary across selection rounds. Preliminary analyses suggested that children's teammate selections were similar for the first and second games, so we did not include this variable in the fixed-effects model. The random-effects model consisted of a random intercept for participant. We used the same weakly informative priors for the fixed effects as in Experiments 1 and 2.

Was there a brilliance bias against girls in children's selections? The main effect of condition was similar in magnitude to the estimates from Experiments 1 and 2, but a substantial portion of its credible interval crossed 0 in this study, $b = -0.225$ [−0.699, 0.246] (for the full model, see Table S5, available in the online supplemental material). However, we observed a credible interaction between condition and selection round, $b = -0.417$ [−0.805, −0.037], such that the effect of the experimental manipulation on children's teammate selections was stronger in later selection rounds (Figure 1). Specifically, it was only on the third selection round that children became less likely to choose girls as teammates for the "smart" game (37.6% [28.6 to 47.4%] girls selected) than for the control game (53.4% [42.8 to 63.7%] girls selected), $b = -0.722$ [−1.497, −0.007].[9] On this round, the odds of selecting a girl (vs. a boy) dropped by 51.4% when the game was said to be for really smart children.

The condition difference on the third selection round was not moderated by participants' gender, $b = -0.614$ [−1.824, 0.602]. That is, girls' bias against girls as teammates for the "smart" (vs. the control) game did not differ credibly from boys' bias against girls. Some caution is warranted is interpreting this result, however: In absolute terms, the condition difference on the third selection round seemed substantially larger for girls than for boys (Figure 2), and it is possible that a study with a larger sample would have detected a credible interaction effect.

Was there ingroup bias in children's responses? Indeed, we observed a credible main effect of gender, $b = 2.36$ [1.88, 2.87], with girls being much more likely to select girls (79.4% [73.0 to 84.7%] girl selections) than boys were (26.6% [20.3 to 33.7%] girl selections). However, the magnitude of this ingroup bias varied across selection rounds, as indicated by a credible interaction between gender and selection round, $b = -0.617$ [−0.991, −0.244]. Although children preferred own-gender teammates in all selection rounds, this preference became considerably weaker in later rounds. For example, girls' and boys' selections of girls as teammates were 60.7 percentage points apart on the first round (88.1 vs. 27.5% girls selected by girls vs. boys, respectively)[10] and only 40.8 percentage points apart on the third round (66.5 vs. 25.7% girls selected by girls vs. boys, respectively).

Putting these two sets of results together, it seems that children start out with a strong ingroup bias that then weakens with each consecutive selection, giving way to a brilliance bias against girls. This interpretation was also supported by an examination of boys' and girls' responses (considered separately) across the three selection rounds (Figure 2). To begin, we note a crucial difference in the expected effect of ingroup and brilliance biases on boys' and girls' responses. Whereas these two biases push boys' responses in the same direction (i.e., toward choosing fewer girls than boys), for girls they point in different directions: An ingroup bias would lead girls to choose more girls, but a brilliance bias would lead them to choose fewer girls (specifically in the experimental condition). Thus, if ingroup bias is relatively more influential in children's early (vs. late) selections and the brilliance bias is relatively more influential in late (vs. early) selections, this pattern should be particularly salient in girls' responses (see Tables S9 and S10, available in the online supplemental material, for the separate models on the girls' and boys' data). Indeed, as illustrated in Figure 2, girls (but not boys) became less likely to select own-gender teammates across selection rounds, $b = -0.629$ [−0.912, −0.359], and this decrease was particularly sharp in the experimental (vs. the control) condition, $b = -0.477$ [−1.024, 0.062], where the game was said to be for children who are "really, really smart."

[9] One may wonder whether the reminder that children received after their first selection in the experimental (but not control) condition explains this difference: Perhaps the reminder caused children to switch their answers from choosing own-gender teammates (ingroup bias) to choosing children of the other gender. Several aspects of the data contradict this possibility. For example, children's selections in the experimental condition were *more* likely to stay the same when they received a reminder vs. when they did not. The percentage of children whose selections were unaffected by the reminder (i.e., children whose first and second selections were either two girls or two boys) was 57.3%. By comparison, only 47.4% of children were consistent (i.e., chose either two girls or two boys) between their second and third selections, which were not separated by a reminder.

[10] These numbers suggest that more girls than boys were chosen in the first round overall (see also Figure 1). This result may be explained by the fact that young girls seem to have a stronger preference for ingroup friends and playmates than boys do (e.g., Shutts et al., 2013; Shutts, Kenward, Falk, Ivegran, & Fawcett, 2017).

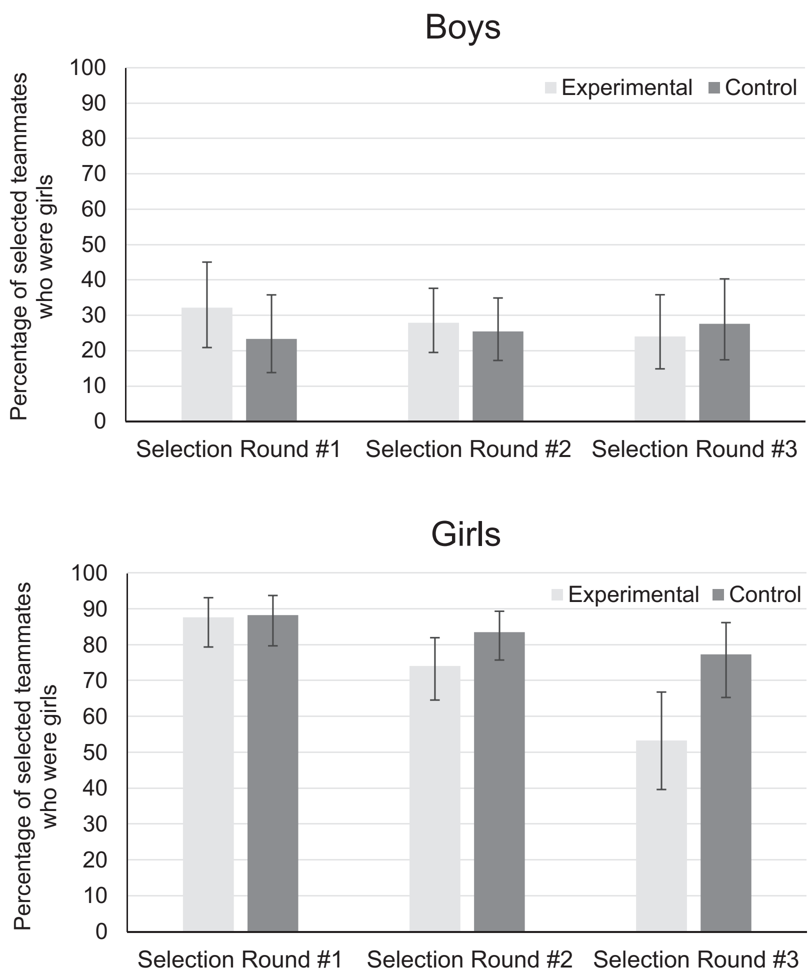


*Figure 2.* The percentage of girls selected as teammates by boys (top) and girls (bottom) in each selection round of Experiment 3, by experimental versus control condition. The error bars represent 95% credible intervals. The percentage estimates and the credible intervals were calculated as marginal effects in Bayesian mixed-effects multilevel logistic regressions.

For boys, the interaction between selection round and condition (experimental vs. control) was in the same direction as for girls but was weaker, $b = -0.306$ [−0.833, 0.235].

It is also noteworthy that, by the third selection round, the credible interval in the experimental condition was entirely below the neutral 50% threshold (28.6 to 47.4% girls selected; Figure 1, right). This result indicates that, as a group, children selected fewer girls than boys for the "smart" game on this round, which suggests the presence of gender bias in the second sense defined above (i.e., a substantial departure from a 50/50 gender split within a condition).[11] In contrast, the credible interval for the control game spanned 50% (42.8 to 63.7% girls selected). The numerical imbalance in children's selections for the "smart" (but not the control) game on this round mirrors the imbalance in adults' referrals for the "brilliance" (but not the control) jobs in Experiments 1 and 2. Another parallel to Experiments 1 and 2 is that the departure from 50/50 balance was driven mostly by the male participants: On the third selection round, boys selected only 24.0% [14.8 to 35.8%] girls for the "smart" game, whereas girls still selected a majority of girls for this game (53.4% [39.8 to 66.9%]). (Recall, however, that girls still showed gender bias in the sense of being less likely to select girls for the "smart" game than for the control game.) The meta-analysis of Experiments 1 and 2 revealed a parallel difference in adults' responses: Men referred only 30.6% [26.4 to 35.1%] women for the "brilliance" job, whereas women referred 54.1% [49.9 to 58.1%] women for this job.

The differences observed in children's behavior across the three selection rounds led us to go back to the adult

[11] However, based on this criterion, children in the experimental condition were biased *in favor of* girls on the first selection round (Figure 1).

data from the first two studies and explore whether adults' first and second referrals differed as well. We used the same data and model as for our meta-analysis above, with the addition of referral round (first or second; Level-1 predictor) and all its interactions with the other variables as fixed effects. This analysis revealed a main effect of referral round, $b = 0.379$ [0.206, 0.558], with a lower probability of female referrals in the first round (42.6% [39.5 to 45.7%]) than in the second round (52.0% [48.9 to 55.1%]). In other words, adult participants were overall more likely to refer a man at first, but their referrals became more gender-neutral on the second round. However, the experimental versus control effect did not differ credibly across the two referral rounds, $b = 0.151$ [−0.197, 0.498]. Thus, adults consistently (i.e., on both rounds) displayed bias against women when making referrals for a brilliance job—a different pattern than that displayed by children, who showed gender bias only in later selection rounds. The difference between adults' and children's response patterns could be due to a variety of factors, especially considering that the two experimental procedures differed in key respects (e.g., referring familiar individuals for a hypothetical job vs. selecting unfamiliar individuals for an ostensibly real team game). For example, one plausible reason for children's ingroup bias in early selection rounds is that children may have thought they would actually have to interact with the children they selected for their team, which might have accentuated the tendency to choose similar others.

Returning to the results of Experiment 3, we observed two other relationships of note. First, older children were less likely than younger children to select a girl, $b = -0.305$ [−0.588, −0.026]. With each additional year of age, a child's odds of selecting a girl for their team decreased by 26.3%. Second, like age, stereotyping was negatively related to the likelihood of selecting a girl, $b = -1.43$ [−2.32, −0.542]. Going from the minimum to the maximum score on the stereotyping measure corresponded to a 76.1% reduction in the odds of choosing a girl as a teammate. The relationship between stereotyping and the outcome did not differ by experimental versus control condition, $b = 0.467$ [−0.946, 1.870], or by children's gender, $b = -0.185$ [−1.595, 1.215].

In summary, young children showed bias against girls in the context of activities said to require high levels of intellectual ability. When asked to select teammates for these activities, children first tended to select other children of their gender (consistent with the strong ingroup bias present at this age; Dunham et al., 2016; Yee & Brown, 1994; Zosuls et al., 2011). However, as they selected more teammates, they increasingly favored boys over girls. No such shift occurred when the experimenter did not comment on the abilities required to succeed on these activities.

# General Discussion

Despite women's achievements in the classroom and the workplace, the three experiments presented here suggest that women may still encounter bias in circumstances where brilliance is viewed as the key to success. We found evidence of such bias in a novel referral paradigm (Experiment 1): Participants were less likely to refer women for a position described as requiring intellectual ability than for a position described as requiring motivation; the gender of the recommender did not seem to matter vis-à-vis this bias. We found the same pattern, although somewhat attenuated, in a preregistered replication of the first experiment with a larger and more diverse sample (Experiment 2). Experiment 3 suggested that this bias is present even in children: When selecting teammates for a game, 5- to 7-year-olds selected teammates of their own gender at first but then chose fewer girls when the game was described as requiring smarts than when it was not.

## Theoretical Contributions

This work makes several theoretical contributions. First, by finding gender bias in contexts where brilliance is valued, our studies add to our current understanding of the processes that lead to women's underrepresentation in "genius fields"—that is, fields such as physics and philosophy, in which success is generally seen as depending on high-level intellectual ability (Leslie, Cimpian, et al., 2015). Prior work suggested that the atmosphere of these "genius fields," and particularly the explicit emphasis on brilliance, has negative effects on women's sense of fitting in with others in these fields (i.e., their sense of belonging) and on their self-efficacy (i.e., the extent to which they expect they will be successful; Bian et al., 2018; see also Good, Rattan, & Dweck, 2012). In turn, these effects take a toll on women's interest as well. Similarly, there is evidence that women are more likely to experience stereotype threat in settings where innate ability is valued (Emerson & Murphy, 2015). In these settings, women expect that others' views of them will be influenced by the negative stereotypes about their abilities; this expectation lowers women's trust and makes them more likely to disengage from the field. The present findings suggest that others' judgments should indeed be reason for concern: A key implication of our studies is that women encounter bias in "genius fields." This bias is thus an important part of any theory seeking to explain why women are underrepresented in these prestigious fields.

Second, these findings add to our understanding of gender bias in professional contexts by examining an early, predecision stage of the recruitment process that has so far not been the target of sustained investigation as a source of gender bias. Even if the hiring *decision* itself is unbiased, if it is based on a pool of candidates in which women are present in fewer numbers than their accomplishments would warrant, a strong case can still be made for bias in the hiring *process*. While gender bias may be becoming less common in employers' and supervisors' "public" behavior (e.g., hiring or promotion decisions; Williams & Ceci, 2015), in part because the possibility of bias is often explicitly discussed in these contexts, young women's path to a successful career goes through many contexts in which people may be less guarded and—our evidence suggests—may still behave in biased ways (see also Nittrouer et al., 2018).

Third, our findings add a developmental dimension to investigations of gender bias. Although it is intuitive to think of gender bias as an adult phenomenon, the gender imbalances currently seen in many academic and professional fields may actually be attributable, in part, to processes that unfold early in development (e.g., Cheryan et al., 2015). Our data support this possibility. After they had a chance to express their ingroup favoritism, the 5- to 7-year-olds in Experiment 3 exhibited gender bias, selecting fewer girls as teammates when the activity was said to be for children who are "really, really smart" than when it was not. This difference was most prominent among the girls in our sample but did not in fact differ credibly by gender. Outside the lab, where the intellectually challenging activities have to do with mathematics, science, chess, and so forth, this sort of biased behavior may deprive girls of opportunities to develop expertise in these domains via interactions with peers. If girls are excluded from the real-world counterparts of our novel activities—or at least not encouraged to engage with them as much as boys are—that could make it more difficult for girls to pursue these domains when they are older. In this way, part of the foundation for women's underrepresentation in "genius fields" may be laid quite early.

## Limitations and Alternative Explanations

One potential limitation of the first two studies is that we cannot know the *actual* intellectual abilities of the referred individuals. This suggests an alternative interpretation: Perhaps our participants happened to know more intellectually gifted men than women, which would mean their referrals were not actually biased. While such differences might emerge by chance in small samples of participants, our combined sample across Experiments 1 and 2 consisted of 1,158 individuals from diverse backgrounds. As groups, the male and female acquaintances of these 1,158 individuals are likely to approximate the characteristics of the general population of men and women, for which there is no compelling evidence of gender differences in "raw" intellectual ability (e.g., Guiso et al., 2008; Spelke, 2005). Of course, the male and female acquaintances of our participants may have differed in other ways that influenced participants' referrals. For example, the men may have had more challenging jobs or occupied positions of higher authority than the women (e.g., Heilman, 2012). These differences could serve as indirect cues to intellectual ability; people often infer others' traits from their social roles (e.g., Cimpian & Salomon, 2014; Koenig & Eagly, 2014). However, insofar as the current occupations of participants' acquaintances may themselves have been shaped by the biases we are examining here, these cues may be misleading and perpetuate gender inequality. We also note that the paradigm used in Experiment 3 is not sensitive to these concerns, insofar as children were asked to select teammates for games from a set of children with whom they had no prior acquaintance. Children's preferential selection of unfamiliar boys over unfamiliar girls as teammates for the "smart" game is a clear example of gender bias (i.e., differential treatment by gender in the absence of specific knowledge of any gender differences in the potential teammates' abilities).

It is also interesting to consider whether participants' referrals in Experiments 1 and 2 were driven by altruistic concerns about the people being referred rather than, as we claim, by gender bias. Perhaps the description of the brilliance job evoked an image of a workplace that is male-dominated and competitive (e.g., Gaucher, Friesen, & Kay, 2011). As a result, participants may have recommended fewer women for this job simply because they were concerned about how they would fit in, not because they harbored negative views about their abilities. Several aspects of our results are inconsistent with this alternative possibility. For instance, the gender of participants' referrals was predicted by their endorsement of negative stereotypes about women's intellectual abilities: The more participants endorsed statements such as, "One is more likely to find a male with a genius-level IQ than a female with a genius-level IQ," the less likely they were to refer a woman, consistent with the gender-bias interpretation of our results. The open-ended reasons participants provided for their referral decisions (see Appendix S1, available in the online supplemental material) support this interpretation as well. We went back to these responses to see whether any included mention of the work environment of the brilliance job, the fit with the other people there, and so forth. We did not see any justifications of this sort—virtually all participants explained their decisions by listing the abilities of the referred individuals and how they matched the job description. Note that, although expectations about the environment of brilliance jobs did not seem to factor into the gender bias observed in these particular studies, this mechanism could operate in other circumstances and may also constitute an obstacle for women hoping to pursue "genius fields."

Another limitation of this research is the use of hypothetical scenarios. Caution is warranted when extrapolating these findings to behaviors outside the lab. Nevertheless, there are also reasons to believe that our experimental paradigms were valid analogs of the relevant real-world situations and thus captured most of the same psychological processes. For instance, participants were asked to refer people they knew, which is precisely what would happen if they had to make a referral in their actual jobs. Moreover, referrals are often made off the cuff in the context of a conversation or in a quick e-mail response, which is not too different from how participants made their referrals in our studies. Similar considerations apply to the task in Experiment 3, which capitalized on the age-appropriate and familiar process of choosing other children with whom to play a game. In sum, though lab studies such as ours have limitations with respect to external validity, other aspects of our methodology mitigate these concerns.

## Open Questions and Directions for Future Work

An important direction for future work would be to identify means of combatting the bias documented here. There are at least two ways to fight this bias: by changing the "brilliance = men" stereotype or by making this stereotype irrelevant to decisions about employment, promotion, awards, and so forth. Given how difficult it is to induce long-lasting change in stereotypes (e.g., Lai et al., 2016), a more effective bias-reduction strategy, at least in the short term, might be to find ways of blocking the impact of the "brilliance = men" stereotype in professional contexts. One could consider, for instance, intervening to change how people talk (or, better yet, think) about success in fields that currently value brilliance. If members of a field believe that success depends on a special quality (brilliance, genius, etc.) that only certain individuals possess and that cannot be acquired (Bian et al., 2018; see also Rattan, Savani, Naidu, & Dweck, 2012), then the logical next step is to determine who among their trainees and peers has this quality. The problem is that there is no predetermined, agreed-upon set of cues to brilliance; these judgments are riddled with ambiguity, and stereotypes and bias thrive on ambiguity (e.g., Koch et al., 2015). One solution might be to step away from murky intuitions about intellectual talent altogether and adopt what is known as a *growth mindset* (e.g., Dweck, 2006), which is the idea that most skills can be considerably improved with focused practice and appropriate mentoring. From the perspective of a growth mindset, an individual's chances of success in a field are not predetermined by the (fixed) amount of talent they possess; rather, what matters most is the individual's willingness to dedicate themselves to learning the relevant skills and content matter, given appropriate guidance. Growth mindsets democratize judgments of "potential," lessening the effect of stereotypes that associate talent with only certain groups (e.g., Yeager et al., 2016).

We should note that it is an open question whether some fields actually require more brilliance than others. To the extent that progress in many fields of academia is increasingly reliant on large collaborative projects, individuals' intellectual abilities may in fact be less important to their success than their ability to work with others in teams (e.g., Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). Nevertheless, the present findings do not speak directly to this issue, and—crucially—our argument holds regardless of whether field-specific beliefs about success are true or false: Whether or not someone needs to be brilliant to do well in "genius fields," if the members of these fields believe this to be the case, their behavior is susceptible to influence by whatever associations their culture attaches to the idea of brilliance and genius (e.g., that it is a male trait).

In the future, this line of research could be extended in several directions. First, it would be useful to have more information about the range of professional contexts in which women encounter bias because of the "brilliance = men" stereotype. We can also ask whether these biases are exacerbated as one moves up the ranks in a field or organization, and whether judgments made at these higher levels are biased in an even broader range of contexts (that is, even beyond fields where brilliance is explicitly valued). Second, we need to know more about how the gender biases documented here intersect with ethnicity. For instance, preliminary evidence suggests that, developmentally at least, the "brilliance = men" association is strongest when evaluating White (vs. Black) men and women (e.g., Boston et al., 2018). Our studies do not speak to these issues: We have no information about the ethnicity of participants' referrals in Experiments 1 and 2, and the potential teammates presented to children in Experiment 3 were all White. It will thus be important to explore how "brilliance = men" biases apply to targets of various ethnicities. Finally, further developmental research is needed to explore the robustness of the brilliance bias we identified among young children and the generalizability of this bias across cultural contexts, as well as to follow up on some of the patterns for which our data did not provide conclusive evidence (e.g., the trend whereby girls seemed to differentiate between the "smart" and the control games more strongly than boys).

## Conclusion

To conclude, our findings reveal an early developing bias against girls and women in circumstances where intellectual ability is thought to be essential for success. Despite the objective evidence of women's educational and professional accomplishments, it seems that their ability to make intellectual contributions is still not seen as being on par with men's. This bias likely represents a major obstacle for women aspiring to prestigious careers in today's society.

## References

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355,* 389–391. http://dx.doi.org/10.1126/science.aah6524

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. http://dx.doi.org/10.1016/j.jesp.2017.11.006

Billington, J., Baron-Cohen, S., & Wheelwright, S. (2007). Cognitive style predicts entry into physical sciences and humanities: Questionnaire and performance tests of empathy and systemizing. *Learning and Individual Differences, 17,* 260–268. http://dx.doi.org/10.1016/j.lindif.2007.02.004

Boston, J. S., Lei, R., Chestnut, E., Vraneski-Shachnai, R., & Cimpian, A. (2018, March). *Children show a robust "brilliance = men" stereotype for white, but not black, targets*. Poster presented at the Illinois Summit on Diversity, Champaign, IL.

Brooks, R. R., & Purdie-Vaughns, V. (2007). Supermodular architecture of inclusion. *Harvard Journal of Law & Gender, 30,* 379–387.

Bürkner, P. C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80,* 1–28.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 3157–3162. http://dx.doi.org/10.1073/pnas.1014871108

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology, 6,* 49. http://dx.doi.org/10.3389/fpsyg.2015.00049

Chestnut, E. K., Lei, R. F., Leslie, S. J., & Cimpian, A. (2018). The myth that only brilliant people are good at math and its implications for diversity. *Education in Science, 8,* 65. http://dx.doi.org/10.3390/educsci8020065

Cimpian, A., & Leslie, S. J. (2015). Women in science: Response to Comment on "Expectations of brilliance underlie gender distributions across academic disciplines." *Science, 349,* 391. http://dx.doi.org/10.1126/science.aaa9892

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. http://dx.doi.org/10.1038/scientificamerican0917-60

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences, 37,* 461–480. http://dx.doi.org/10.1017/S0140525X13002197

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7–29. http://dx.doi.org/10.1177/0956797613504966

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review, 25,* 207–218. http://dx.doi.org/10.3758/s13423-017-1266-z

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science, 19,* 781–789. http://dx.doi.org/10.1111/desc.12321

Dutt, K., Pfaff, D. L., Bernstein, A. F., Dillard, J. S., & Block, C. J. (2016). Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience, 9,* 805–808. http://dx.doi.org/10.1038/ngeo2819

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

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

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,* 735–754. http://dx.doi.org/10.1037/0022-3514.46.4.735

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly, 18,* 585–609. http://dx.doi.org/10.1111/j.1471-6402.1994.tb01049.x

Emerson, K. T., & Murphy, M. C. (2015). A company I can trust? Organizational lay theories moderate stereotype threat for women. *Personality and Social Psychology Bulletin, 41,* 295–307. http://dx.doi.org/10.1177/0146167214564969

Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles, 30,* 81–92. http://dx.doi.org/10.1007/BF01420741

Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood. *Journal of Personality and Social Psychology, 97,* 517–532. http://dx.doi.org/10.1037/a0016030

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,* 878–902. http://dx.doi.org/10.1037/0022-3514.82.6.878

Fyock, J., & Stangor, C. (1994). The role of memory biases in stereotype maintenance. *British Journal of Social Psychology, 33,* 331–343. http://dx.doi.org/10.1111/j.2044-8309.1994.tb01029.x

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology, 101,* 109–128. http://dx.doi.org/10.1037/a0022530

Gërxhani, K., & Koster, F. (2015). Making the right move: Investigating employers' recruitment strategies. *Personnel Review, 44,* 781–800. http://dx.doi.org/10.1108/PR-12-2013-0229

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102,* 700–717. http://dx.doi.org/10.1037/a0026659

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Diversity. Culture, gender, and math. *Science, 320,* 1164–1165. http://dx.doi.org/10.1126/science.1154094

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,* 153–166. http://dx.doi.org/10.1007/s11199-011-9996-2

Halim, M. L. D., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., & Amodio, D. M. (2017). Gender attitudes in early childhood: Behavioral consequences and cognitive antecedents. *Child Development, 88,* 882–899. http://dx.doi.org/10.1111/cdev.12642

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269,* 41–45. http://dx.doi.org/10.1126/science.7604277

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior, 32,* 113–135. http://dx.doi.org/10.1016/j.riob.2012.11.003

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48,* 1393–1409. http://dx.doi.org/10.3758/s13428-015-0647-3

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592. http://dx.doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Diversity. Gender similarities characterize math performance. *Science, 321,* 494–495. http://dx.doi.org/10.1126/science.1160364

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology, 100,* 128–161. http://dx.doi.org/10.1037/a0036734

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. http://dx.doi.org/10.1037/a0037215

Kolker, A. F. (2017). Gender segregation in everyday life. In B. S. Turner (Ed.), *The Wiley Blackwell encyclopedia of social theory* (pp. 1–3). New York, NY: Wiley. http://dx.doi.org/10.1002/9781118430873.est0460

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25,* 178–206. http://dx.doi.org/10.3758/s13423-016-1221-4

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., . . . Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General, 145,* 1001–1016. http://dx.doi.org/10.1037/xge0000179

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

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open, 6,* 1–17. http://dx.doi.org/10.1177/2158244016636433

Magua, W., Zhu, X., Bhattacharya, A., Filut, A., Potvien, A., Leatherberry, R., . . . Kaatz, A. (2017). Are female applicants disadvantaged in National Institutes of Health peer review? Combining algorithmic text mining and qualitative methods to detect evaluative differences in R01 reviewers' critiques. *Journal of Women's Health, 26,* 560–570. http://dx.doi.org/10.1089/jwh.2016.6021

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. http://dx.doi.org/10.3389/fpsyg.2015.00235

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23,* 103–123. http://dx.doi.org/10.3758/s13423-015-0947-8

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 16474–16479. http://dx.doi.org/10.1073/pnas.1211286109

National Association for Gifted Children. (2015). *2014–2015 state of the states in gifted education: Policy and practice data* [Data file]. Retrieved from https://bit.ly/2vPif0Y

National Center for Education Statistics. (2017). *Bachelor's, master's, and doctor's degrees conferred by postsecondary institutions, by sex of student and discipline division: 2014–15*. Retrieved from https://bit.ly/2MrmiYz

Nittrouer, C. L., Hebl, M. R., Ashburn-Nardo, L., Trump-Steele, R. C. E., Lane, D. M., & Valian, V. (2018). Gender disparities in colloquium speakers at top universities. *Proceedings of the National Academy of Sciences of the United States of America, 115,* 104–108. http://dx.doi.org/10.1073/pnas.1708414115

Pew Research Center. (2015). *Women and leadership: Public says women are equally qualified, but barriers persist*. Washington, DC: Author. https://pewrsr.ch/2niKcdO

Powlishta, K. K. (1995). Gender bias in children's perceptions of personality traits. *Sex Roles, 32,* 17–28. http://dx.doi.org/10.1007/BF01544755

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,* 787–803. http://dx.doi.org/10.1037/a0029263

Shutts, K., Kenward, B., Falk, H., Ivegran, A., & Fawcett, C. (2017). Early preschool environments and gender: Effects of gender pedagogy in Sweden. *Journal of Experimental Child Psychology, 162,* 1–17. http://dx.doi.org/10.1016/j.jecp.2017.04.014

Shutts, K., Pemberton, C. K., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development, 14,* 35–62. http://dx.doi.org/10.1080/15248372.2011.638686

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60,* 950–958. http://dx.doi.org/10.1037/0003-066X.60.9.950

Stan Development Team. (2017a). *Prior choice recommendations*. Retrieved from https://bit.ly/2bGo1d4

Stan Development Team. (2017b). *Stan modeling language: User's guide and reference manual*. Retrieved from https://bit.ly/2Ok9CmZ

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles, 41,* 509–528. http://dx.doi.org/10.1023/A:1018839203698

Stephens-Davidowitz, S. (2014, January 18). Google, tell me. Is my son a genius? *The New York Times*. Retrieved from https://nyti.ms/1azD4xD

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,* e0150194. http://dx.doi.org/10.1371/journal.pone.0150194

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods, 22,* 217–239. http://dx.doi.org/10.1037/met0000100

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140,* 1174–1204. http://dx.doi.org/10.1037/a0036620

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25,* 35–57. http://dx.doi.org/10.3758/s13423-017-1343-3

Wenneras, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature, 387,* 341–343. http://dx.doi.org/10.1038/387341a0

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 5360–5365. http://dx.doi.org/10.1073/pnas.1418878112

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330,* 686–688. http://dx.doi.org/10.1126/science.1193147

Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., . . . Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences of the United States of America, 113*(24), E3341–E3348. http://dx.doi.org/10.1073/pnas.1524360113

Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology, 33,* 183–196. http://dx.doi.org/10.1111/j.2044-8309.1994.tb01017.x

Zosuls, K. M., Martin, C. L., Ruble, D. N., Miller, C. F., Gaertner, B. M., England, D. E., & Hill, A. P. (2011). "It's not that we hate you": Understanding children's gender attitudes and expectancies about peer relationships. *British Journal of Developmental Psychology, 29,* 288–304. http://dx.doi.org/10.1111/j.2044-835X.2010.02023.x

Zottoli, M. A., & Wanous, J. P. (2000). Recruitment source research: Current status and future directions. *Human Resource Management Review, 10,* 353–382. http://dx.doi.org/10.1016/S1053-4822(00)00032-2

Received April 27, 2018
Revision received October 14, 2018
Accepted October 15, 2018 ■