ORIGINAL ARTICLE

Gender stereotypes and incremental beliefs in STEM and non-STEM students in three countries: relationships with performance in cognitive tasks

Angelica Moè[1] · Markus Hausmann[2] · Marco Hirnstein[3]

Abstract

Women's underrepresentation in science, technology, engineering and mathematics (STEM) has been linked, among others, to gender stereotypes and ability-related beliefs as well as gender differences in specific cognitive abilities. However, the bulk of studies focused on gender stereotypes related to mathematics. The present study, therefore, aimed to map gender stereotypes and incremental beliefs (i.e., the conviction about modifiability) with respect to a wide range of stereotypical male-favouring and female-favouring abilities. Gender stereotypes and incremental beliefs were assessed with self-report questionnaires in 132 STEM students (65 women) and 124 non-STEM students (73 women) in three European countries ranked in the top, middle, and bottom of the Global Gender Gap Report. Moreover, a mental rotation and a verbal fluency test were completed. Men endorsed male-favouring stereotypes more than women, and women endorsed female-favouring stereotypes more than men, an effect that was most pronounced in the country with the larger gender gap. Male STEM students endorsed male-favouring stereotypes more strongly than male non-STEM and female STEM students. Male non-STEM students endorsed female-favouring stereotypes less than female and male STEM students. Female STEM students reported higher incremental beliefs than female non-STEM students, especially in the country with the lowest gender gap. Men outperformed women, and STEM students outperformed non-STEM in mental rotation, while women outperformed men in verbal fluency. Male STEM students' stronger endorsement of male-favouring stereotypes might reflect genuine group differences, at least in mental rotation. While potentially such gender stereotypes can help creating a "chilly climate" where women in academic STEM degrees are expected to perform poorly, those women believed more in the possibility to change and improve in male-favouring abilities which could help them to overcome the potential negative effect of stereotyping.

Introduction

Disproportionately few women decide to study or work in science, technology, engineering, and mathematics (STEM) fields (National Science Foundation 2018). Research identified a number of relevant factors that contribute to women's underrepresentation, for example, performance in gatekeeper tests (Ceci, Williams, & Barnett 2009), wider career choice in women than men due to higher proficiency in both verbal and mathematical skills (Wang, Eccles, & Kenny, 2013), misidentification with mathematics (Nosek, Banaji & Greenwald, 2002), hostility towards women in STEM degrees (the so-called "chilly climate": Seymour & Hewitt, 1997), gender differences in specific cognitive tasks (Levine, Foley, Lourenco, Ehrich, & Ratliff, 2016), and also gender stereotypes and ability-related beliefs (Ceci 2017; Ceci, Ginther, Kahn, & Williams 2014). The current study will focus on endorsed gender stereotypes, beliefs about modifiability, and cognitive performance in two gender-sensitive tasks, with the aim to compare them in female and male STEM and non-STEM students in three Western European countries with different levels of gender equality.

✉ Angelica Moè
angelica.moe@unipd.it

Markus Hausmann
markus.hausmann@durham.ac.uk

Marco Hirnstein
marco.hirnstein@uib.no

[1] Dipartimento di Psicologia Generale, Università degli Studi di Padova, Via Venezia, 8, 35131 Padua, Italy

[2] Department of Psychology, Durham University, Durham, UK

[3] Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway

Gender stereotypes and ability-related beliefs

Already in the first years of primary school girls self-report to be less able than boys in mathematics (Fredericks & Eccles 2002; Moè 2018a) and identify with mathematics less than boys (Cvencek, Meltzoff, & Greenwald 2011). A range of biological, experiential, and motivational factors play a role in shaping these beliefs. For instance, prenatal androgen hormone exposure, which differs between genders, relates with math performance (Bull & Benson, 2006). Probably due to differences in right parietal lobe activation, men tend to rely more than women on a spatial representation of numbers, which favour basic number processing (Bull, Cleland, & Mitchell, 2013), and increase speed in a number comparison task (Huber, Nuerk, Reips, & Soltanlou, 2017). Further confirmation of these gender differences comes from studies considering embodied cognition. For instance, Lugli, D'Ascenzo, Borghi and Nicoletti (2018) found that clockwise movement favours math performance in men but not in women, confirming again that men rely more than women on a spatial number representation. Furthermore, teachers (Li, 1999) and parents (Tomasetto, Alparone, & Cadinu, 2011) believe boys to be more skilled than girls in mathematics. These negative stereotypes about mathematics might prompt girls to engage less in mathematics, science, and technical subjects, which in turn leads to preferences for academic degrees and professions with low math or spatial content (Eddy & Brownell 2016). Interestingly, the magnitude of the gender stereotype related to mathematical skills appeared to be modulated by academic degree: female STEM students were reported to hold fewer implicit gender stereotypes than female non-STEM students (Nosek & Smyth 2011; Smeding 2012). Moreover, also the relationship between gender stereotypes and performance was modulated by academic degree: math grades correlated negatively with gender stereotypes only for female non-STEM students, not for men and female STEM students (Smeding 2012), suggesting that experience and success in science and math fields might help reducing the expectation to fail and favour effort display, performance, and the choice of STEM careers.

As noted by Ceci (2017), not only gender stereotypes but also ability-related beliefs appear to be relevant for women's underrepresentation in STEM fields. The so-called "incremental beliefs" are an example of such ability-related beliefs. Incremental beliefs refer to the personal conviction that abilities are not fixed entities but can be improved with practice, exercise, experience, effort, or more learning (Dweck 1999). Gender stereotypes and incremental beliefs are typically negatively correlated, that is, the stronger the gender stereotype about a specific ability, the less individuals are convinced this ability can be improved through training (Levy, Stroessner, & Dweck 1998). Two studies found an increase

in women's mental rotation performance, when it was highlighted that context influences gender differences more than genetic factors (Moè, 2012) or when women received training where they were told that abilities are modifiable through effort, exercise, and using the most effective strategies (Moè 2016a). Moreover, the stronger women's belief that it is possible to improve in tasks in which men usually excel, the better the performance in mental rotation (Moè, Meneghetti, & Cadinu 2009). In general, participants prefer tasks and school subjects in which they consider themselves to be skilled and if they think improvement is likely (Leslie, Cimpian, Meyer, & Freeland 2015). Thus, we could speculate that men and STEM students hold stronger incremental beliefs than women and non-STEM students about male-favouring abilities and that women and STEM students have similar beliefs about female-favouring abilities. So far, however, incremental beliefs beyond mathematics were hardly investigated in STEM or non-STEM students.

There are gender stereotypes and ability-related beliefs beyond mathematics that seem relevant for women's underrepresentation in STEM fields. However, much less is known about their prevalence, magnitude, and developmental trajectory. For example, a number of studies found that men and women were stereotypically believed to be better at spatial and verbal abilities, respectively (Halpern & Tan 2001; Hausmann, Schoofs, Rosenthal, & Jordan 2009; Hausmann 2014; Hirnstein, Coloma Andrews, & Hausmann 2014; Moè, Meneghetti, & Cadinu 2009). Already 10- to 12-year-olds believed that boys/men have better spatial (Vander Heyden, van Atteveldt, Huizinga, & Jolles 2016) and girls/women better verbal skills (Kurtz-Costes, Copping, Rowley, & Kinlaw 2014). Most of the studies on gender stereotypes in spatial and verbal abilities tested psychology students or did not specify the academic background of their sample. Only a few studies directly tested STEM and non-STEM students, despite the relevance that these gender stereotypes might have for choosing STEM degrees. Hausmann (2014) found that the male/spatial stereotype was more pronounced in STEM students (relative to non-STEM students), while the female/verbal stereotype was stronger in non-STEM students as compared to STEM students.

Gender differences in cognitive tasks

Boys/men typically outperform girls/women in mental rotation (Halpern 2012; Peters, Lehmann, Takahira, Takeuchi, & Jordan 2006), the ability to mentally maintain, manipulate, and rotate 2-D or 3-D figures in space accurately and rapidly. Mental rotation is widely considered to be the strongest cognitive gender difference with effect sizes around $d = 0.8$ (Linn & Petersen 1985; Voyer, Voyer, & Bryden 1995; Zell, Krizan, & Teeter 2015). Many mental rotation tests (e.g., Peters et al., 1995) use abstract 3-D cube figures developed

by Shepard & Metzler (1971), but the male advantage also emerged with other type of stimuli (Voyer & Jansen 2016). Meta-analyses (Linn & Petersen 1985; Voyer et al. 1995) also found a male advantage in other spatial tasks, such as the Water Level Test, in which participants draw a line in tilted bottles to indicate the (horizontal) water orientation, or the Paper Folding Test, where participants were asked to imagine what cube figures that are flattened out would look like if folded, but the effect sizes were in the small-to-medium range.

In turn, a reliable female advantage emerged in verbal memory (Halpern 2012; Miller & Halpern 2014; Andreano & Cahill 2009) and reading comprehension (Reilly 2012; Stoet & Geary 2013). The arguably most researched cognitive gender difference favouring women, however, is verbal fluency (e.g., Scheuringer, Wittig, & Pletzer 2017; Hausmann et al. 2009; Hyde & Linn 1988). In verbal fluency tasks, participants are typically instructed to generate as many words as possible that fulfil a certain semantic criterion (e.g., naming animals, fruits, or things that are red) or a certain linguistic criterion (e.g., words that begin with a specific letter) (Lezak, Howieson, Bigler, & Tranel 2012). The effect size was estimated $d = 0.33$, according to a meta-analysis by Hyde and Linn (1988). More recent studies (Halari, Hines, Kumari, Mehrotra, Wheeler, Ng, & Sharma 2005; Hausmann et al., 2009; Herlitz, Airaksinen, & Nordström 1999; Hirnstein et al. 2014; Hirnstein, Freund, & Hausmann 2012) were roughly in line with this number.

Again, only a few studies specifically compared STEM and non-STEM students' performances (e.g., Moè, Jansen & Pietsch 2018), as the majority of studies tested psychology students or students whose degree was not specified. The available data corroborated the view that STEM students outperform non-STEM in mental rotation (e.g., Hausmann 2014; Moè 2016b; Moè et al. 2018; Peters, Laeng, Latham, Jackson, Zaiyouna, & Richardson 1995; Sanchis Segura, Aguirre, Cruz-Gómez, Solozano, & Forn 2018) and there was tentative support that non-STEM students outperform STEM students in verbal fluency (e.g., North 2005; Hausmann 2014). Moreover, Hausmann (2014) found that men and STEM students had higher mental rotation scores than women and non-STEM students, respectively. Female STEM students had particularly low mental rotation scores when primed with gender stereotypes. In verbal fluency, women outperformed men only when gender stereotypes were not primed, regardless of the participants' academic background. Sanchis Segura et al. (2018) found that male non-STEM students outperformed female non-STEM students only when the instructions implied that men perform better. These findings are in line with a bulk of research, showing that priming gender stereotypes can reduce or boost mental rotation and verbal fluency performance in both genders (e.g., Heil, Jansen, Quaiser-Pohl, & Neuburger 2012; Moè

& Pazzaglia 2006; Moè 2009; Wraga, Duncan, Jacobs, Helt, & Church 2006).

Moreover, gender gaps regarding economic participation, educational attainment, health and survival, as well as political empowerment (World Economic Forum, 2016, 2018) could play a role in maintaining gender differences in gender stereotypes endorsement, incremental beliefs, and objective scores. To the best of our knowledge, this is the first attempt to include the factor gender gap/country as a further explanation for the observed differences in gender stereotypes, ability-related beliefs, and cognitive performances.

Aims and hypotheses

The present study had four goals. First, we aimed to measure the magnitude of explicit gender stereotypes in male and female STEM and non-STEM students with respect to a wide range of stereotypical male-favouring and female-favouring abilities. Based on findings in mathematics and implicit stereotype assessments (e.g., Cvencek, Meltzoff, & Greenwald 2011), we hypothesized that male students endorse male-favouring stereotypes more strongly than female students, and female students endorse female-favouring stereotypes more strongly than male students. Moreover, we expected female STEM students to have less pronounced male-favouring stereotypes than female non-STEM students and male non-STEM students to have less pronounced female-favouring stereotypes than male STEM students, based on similar findings in mathematics (Nosek & Smyth 2011; Smeding 2012). Second, we aimed to map "incremental beliefs" in STEM and non-STEM students with respect to those gender stereotypes and hypothesized that female STEM and male non-STEM students would display higher incremental beliefs in male-favouring and female-favouring abilities, respectively (implying stronger conviction that those abilities can be improved through training and effort). Third, we investigated cognitive tasks that were frequently shown to reveal gender differences and expected to find better mental rotation performance in men and STEM students (as compared to women and non-STEM students), and better verbal fluency performance in women and non-STEM students (as compared to men and STEM students), based on findings by Voyer et al. (1995), Hyde and Linn (1988), North (2005) and Hausmann (2014). In addition, in an exploratory analysis, we aimed to assess whether the students' gender stereotypes and incremental beliefs correlated with their cognitive performance. Finally, since gender stereotypes can vary across countries (Szameitat, Hamaida, Tulley, Saylik, & Otermans 2015), and consequently maybe also the ability-related beliefs and cognitive performances,

we included participants from three different European countries ranked in the upper, middle, and lower sections of the Global Gender Gap Report (World Economic Forum, 2016, 2018) to directly test the hypothesis that the higher the gender gap, the higher the endorsed stereotypes, and the lower the incremental beliefs and cognitive performance.

Methods

Participants

In total, 120 men and 138 women ($M = 20.91$ years of age, SD = 2.09) participated on a voluntary basis, recruited at three universities (Padua, Durham, Bergen) in three different Western European countries (i.e., Norway, United Kingdom, and Italy). According to the Global Gender Gap Report-Western Europe section at the time of data collection and now (World Economic Forum, 2016; 2018), which comprised 20 countries, Norway, United Kingdom, and Italy ranked in the upper (3rd, 2016; 2nd, 2018), middle (11th, 2016; 9th, 2018), and lower sections of the ranking (16th, 2016; 17th, 2018), respectively. Lower rankings represent larger gender equality gaps.

Two participants were excluded, because they were registered for a combined honour degree in Psychology and Theology, which did not allow explicit allocation to STEM or non-STEM disciplines. Of the remaining 256 participants, 132 students (65 women, 67 men) were taking single honour STEM degrees in either Mathematics, Physics, Engineering, Chemistry, or Sciences, and 124 students (73 women and 51 men) were enrolled as single honour non-STEM students including Languages, Education, Philosophy, and History (see Table 1, for an overview). A post hoc power analysis using G*power (Faul, Erdfelder, Lang, & Buchner 2007) revealed that to detect an effect with $f = 0.25$ (medium-effect size) for a three-way interaction the sample provided reasonable power (0.93), given the factorial design of 2 (gender) × 2 (academic degree) × 3 (country), $p < 0.05$, numerator $df = 3$, groups = 12.

Table 1 Overview number of participants

	Global gender gap—Western Europe section			
	High	Medium	Low	Total
Women				
STEM	24	20	21	65
Non-STEM	28	25	20	73
Men				
STEM	21	24	22	67
Non-STEM	17	15	19	51
Total	90	84	82	256

Table 2 Mean endorsed gender stereotypes scores and factor loadings of the beliefs questionnaire (part 1)

Item	Mean	Factor 1 Male-favouring	Factor 2 Female-favouring
Drawing	− 0.54***		0.46
Solving math problems	0.74***	0.60	
Learning a foreign language	− 0.69***		0.55
Solving a puzzle	0.18**	0.51	0.35
Being curious to learn new things	− 0.17**		0.43
Paying attention during a lecture	− 0.81***		0.45
Feeling others' emotions	− 1.60***		0.70
Engaging in a new sport	1.04***	0.53	
Recalling people's names	− 0.63***		0.36
Finding a route on a map	0.75***	0.65	
Make calculations in mind	0.72***	0.69	
Finding the quickest way to reach a place	0.65***	0.71	
Being creative	− 0.59***		0.55
Building or repairing something	1.38***	0.64	
Reasoning in a logical and sequential manner	0.61***	0.60	

Mean values can range from − 3 = "women definitely better" to + 3 = "men definitely better", with 0 = "men and women equal". Asterisks indicate significant deviations from 0 using one-sample t tests

***$p < 0.001$; **$p < 0.01$

Measures

Gender stereotypes and incremental beliefs

We employed the Beliefs questionnaire (Moè et al. 2009). It lists 15 abilities such as "drawing", "solving math problems", or "being creative" (for a complete list see Table 2). Participants were presented the same list twice on two different sheets. On the first sheet (gender stereotypes part), participants were asked to rate "How much do you think men and women differ in the following abilities?", on a scale from -3 ("definitely women better") to 3 ("definitively men better"). The score of "0" coded for "men and women equal". Stronger negative scores and stronger positive scores would thus indicate stronger female and male-favouring stereotypes, respectively. On the second sheet (incremental beliefs), students were given the same items but with a different instruction: "Think now how much each of these abilities is modifiable". Then, they rated each on a seven-point Likert scale, anchoring points 1 ("not at all") and 7 ("very much"). The higher the scores, the higher the incremental beliefs. See the section "Results" for reliability.

Mental rotation

We used the Redrawn Vandenberg and Kuse Mental Rotation Test, Version A (MRT) (Peters et al. 1995) that was originally developed by Vandenberg and Kuse (1978) and consists of drawings of 3-dimensional cube figures (Shepard & Metzler 1971). This test, which is the most frequently used paper–pencil test to assess mental rotation abilities, presents 24 items made of five Shepard–Metzler figures: one target and four sample figures. Two of these four sample figures are identical but rotated versions of the target figure. Participants are asked to identify those two identical but rotated figures. For scoring, one point was given if both identical figures were correctly identified. The maximum score was 24.

Verbal fluency

A standard Verbal Fluency test (the VF from the Leistungsprüfsystem: LPS, Horn 1962) was used where participants were asked to write down as many words as possible that start with the letters "F", "A", and "S". For each letter, participants had 1 min. For scoring, the number of words written was computed, excluding those with the same root (e.g., fish, fisher was one point) or proper names.

Procedure

Participants were tested individually in a quiet place. After having signed an informed consent, they were asked to

perform the MRT (3 min for the first 12-items, 3 min break, and 3 min for items 13–24) and the VF tests (in counterbalanced order), followed by the Beliefs questionnaire (first the gender stereotypes items, then the incremental beliefs items) and questions regarding demographics. Afterwards, participants were thanked and debriefed.

Data analysis

To examine whether participants endorse gender stereotypes, we first ran one-sample t tests with the test score 0 (representing no differences in ratings for men and women) for each of the 15 items. Subsequently, we conducted a factor analysis following the recommendations of Neill (2008), to obtain a composite score for further analysis. To assess the hypothesis that gender stereotypes, incremental beliefs, and cognitive performances vary across gender and academic degrees, a series of $2 \times 2 \times 3$ univariate ANOVAs were run with gender, degree, and country (from the three universities) as between-participants factors. Effect sizes are provided as Cohen's d to ease comparisons with previous studies. According to Cohen (1988), values between 0.20 and 0.49, 0.50 and 0.79, and 0.80 and higher are considered small, medium, and large, respectively. One-tailed, post hoc comparisons were conducted with Bonferroni adjustment. The alpha level was set to $p = 0.05$, if not stated otherwise. To assess relationships between performance in the two cognitive tasks, gender stereotypes, and incremental beliefs, a series of Pearson's correlations was run.

Results

Gender stereotypes and incremental beliefs

The means of the 15 items referring to endorsed gender stereotypes are shown in Table 2. All differed significantly from the midpoint of zero, as indicated by one-sample t tests, suggesting the presence of gender stereotypes. The data screening showed that with 256 participants for 15 items, we had a satisfactory participant-to-item ratio of approximately 17:1. A number of indicators were checked to assess overall suitability for a factor analysis. First, the determinant derived from the correlation matrix was 0.049 and thus above the recommended value of 0.00001. Moreover, inter-correlations were well below $r = 0.80$, suggesting that there was no multicollinearity. Second, 12 out of 15 items correlated with at least one other item $r \geq 0.30$. Third, the Kaiser–Meyer–Olkin measure of sampling adequacy (0.73) was above 0.60, Bartlett's test of sphericity was significant $[\chi^2(105) = 751, p \leq 0.001]$, and all communalities were above 0.30. Based on these indicators, the data including all 15 items were regarded suitable for factor analysis.

We carried out a principal component analysis with direct oblimin rotation, since we expected that the underlying factors are correlated. Five factors had eigenvalues greater than 1, the first two explaining 23% and 13% of the variance, respectively, while the last three factors explained 8% variance each. A two-factor solution seemed most appropriate: first, the characteristic bend in the scree plot as reflected by the eigenvalues occurred after two factors. Second, we compared the observed eigenvalues to randomly generated eigenvalues based on 15 variables, 256 participants, and 100 replications with the tool "Monte Carlo PCA for parallel analysis" (Watkins 2000). Only the eigenvalues of the first two observed factors (3.4 and 1.9) were above the randomly generated eigenvalues (1.4 and 1.3), while subsequent observed eigenvalues were level with or below the randomly generated ones.

We then re-ran the principal component analysis with the two-factor solution preselected, explaining a total variance of 36%. Factor loadings higher than 0.30 are presented in Table 2. As can be seen, factor 1 represents items that were rated as favouring males, while factor 2 only contained items where females were rated better. The item "solving a puzzle" loaded on both factors and was, therefore, discarded. In a last step, we analysed the reliability for the seven male- and female-favouring items. Cronbach's alpha for the male-favouring stereotypes was an acceptable 0.77, while for female-favouring stereotypes, it was 0.57, suggesting that subsequent results regarding female-favouring stereotypes must be interpreted with caution.

Differences between gender and academic degree

Gender stereotypes

We then computed a mean composite score across the seven items that loaded on male-favouring and female-favouring stereotypes for each individual and subjected them to the aforementioned ANOVAs. As expected, the intercept with the grand mean of $M = 0.84$ (SD = 0.65) and $M = -0.72$ (SD = 0.52) for male- and female-favouring gender stereotypes, respectively, deviated significantly from zero, $F(1, 244) = 486.86$, $p < 0.001$, $\eta_p^2 = 0.67$, and $F(1, 244) = 514.63$, $p < 0.001$, $\eta_p^2 = 0.68$, confirming that across all participants, these abilities were considered male-favouring and female-favouring, respectively.

A significant main effect gender, $F(1,244) = 6.16$, $p = 0.014$, $\eta_p^2 = 0.03$, and $F(1,244) = 15.89$, $p < 0.001$, $\eta_p^2 = 0.06$, for male-favouring and female-favouring stereotypes, respectively, showed that men endorsed male-favouring stereotypes ($M = 0.96$, SD = 0.71) more than women ($M = 0.74$, SD = 0.58), $d = 0.34$, and that women ($M = -0.83$, SD = 0.56) endorsed female-favouring stereotypes more than men ($M = -0.59$, SD = 0.43), $d = 0.48$. A significant

interaction between gender and country, $F(2, 244) = 9.89$, $p < 0.001$, $\eta_p^2 = 0.08$, and $F(2, 244) = 3.33$, $p = 0.038$, $\eta_p^2 = 0.03$, for male- and female-favouring stereotypes, respectively, revealed that only for the Italian sample men endorsed male-favouring stereotypes significantly more than women with $d = 0.91$ [$t(88) = 4.29$, $p < 0.001$], and women endorsed female-favouring stereotypes more than men with $d = 0.75$ [$t(88) = 3.51$, $p = 0.001$]. There were no significant gender differences in the UK and Norway samples. As a result, a significant main effect country emerged for the male-favouring stereotypes, $F(2, 244) = 7.08$, $p = 0.001$. $\eta_p^2 = 0.06$, and for the female-favouring stereotypes, $F(2, 244) = 3.48$, $p = 0.032$, $\eta_p^2 = 0.03$, yielding that Italian students endorsed male-favouring and female-favouring stereotypes more than UK and Norwegian students (see Table 3).

Moreover, the interaction between gender and academic degree was significant for both male-favouring and female-favouring gender stereotypes, $F(1, 244) = 9.43$, $p = 0.002$, $\eta_p^2 = 0.04$, and $F(1, 244) = 8.28$, $p = 0.004$, $\eta_p^2 = 0.03$, respectively. Male STEM students ($M = 1.09$, SD = 0.71) endorsed male-favouring stereotypes significantly more than female STEM students ($M = 0.67$, SD = 0.56), [$t(116) = 2.46$, $p = 0.008$, $d = 0.66$], and more than male non-STEM students [$t(116)2.46$, $p = 0.008$, $d = 0.45$], while there was no difference between male ($M = 0.78$, SD = 0.66) and female non-STEM students ($M = 0.81$, SD = 0.60, $p > 0.05$). Male non-STEM students ($M = -0.47$, SD = 0.38) endorsed female-favouring stereotypes significantly less than female non-STEM students ($M = -0.90$, SD = 0.55), [$t(122) = 4.82$, $p < 0.001$, $d = 0.88$], and less than male STEM students ($M = -0.67$, SD = 0.45), [$t(116) = 2.61$, $p = 0.010$, $d = 0.47$], while there was no difference between male ($M = -0.67$, SD = 0.45) and female STEM students ($M = -0.75$, SD = 0.56, $p > 0.05$), see Fig. 1a, b. None of the other main

Table 3 Mean endorsed gender stereotypes in the three countries (in brackets SDs)

	Male-favouring	Female-favouring
Italy		
Males	1.38 (0.82)	−0.59 (0.55)
Females	0.73 (0.62)	−1.04 (0.64)
Overall	1.01 (0.78)	−0.85 (0.64)
UK		
Males	0.82 (0.63)	−0.60 (0.43)
Females	0.67 (0.52)	−0.66 (0.47)
Overall	0.74 (0.57)	−0.63 (0.45)
Norway		
Males	0.69 (0.46)	−0.57 (0.28)
Females	0.84 (0.60)	−0.74 (0.45)
Overall	0.77 (0.54)	−0.65 (0.38)
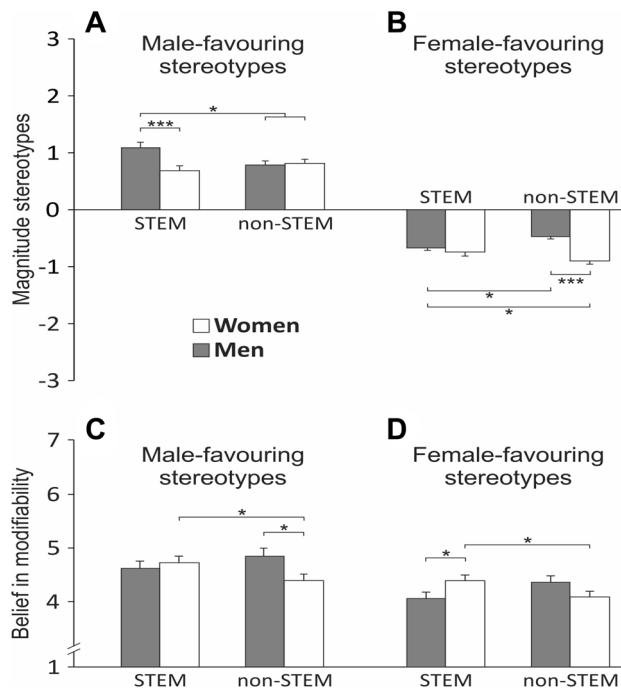
Fig. 1 Mean male-favouring gender stereotypes (bars indicate SEM) and incremental beliefs about male-favouring (**a**, **c**) and female-favouring (**b**, **d**) abilities. Men taking STEM degrees endorse male-favouring gender stereotypes more than male non-STEM students and women. Men taking a non-STEM degree endorse female-favouring gender stereotypes less than female STEM students and less than male STEM. Women pursuing an STEM degree believe more in the modifiability of typical male-favouring or female-favouring abilities than women in non-STEM degrees. For gender beliefs, 0=no gender difference, positive scores=men better, negative scores=women better (range -3="definitely better women" to 3="definitely better men"). For incremental beliefs, 1="not at all" and 7="very much" when asked "Think now how much each of these abilities is modifiable". ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

effects or interactions were significant (all Fs ≤ 1.97, all ps ≥ 0.141).

Incremental beliefs

From the incremental beliefs part of the beliefs questionnaire, we extracted those seven items that corresponded to male- and female-favouring gender stereotypes (see Table 2) and computed one mean each for male-favouring (Cronbach's alpha=0.75) and female-favouring (Cronbach's alpha=0.75) abilities. Higher values reflect stronger incremental beliefs. Incremental belief scores were subjected to the $2 \times 2 \times 3$ ANOVAs, as described above.

The ANOVAs revealed a significant interaction gender by academic degree (male-favouring: $F(1, 244) = 5.63$, $p = 0.018$, $\eta_p^2 = 0.02$; female-favouring: $F(1, 244) = 5.51$, $p = 0.020$, $\eta_p^2 = 0.02$). Female STEM students rated the

possibility to improve in male-favouring ($M = 4.71$, SD = 1.01) domains similarly as male STEM students ($M = 4.61$, SD = 1.11) and more than female non-STEM students ($M = 4.39$, SD = 0.93), $t(136) = 1.92$, $p = 0.029$, $d = 0.33$. Those female non-STEM students had, in turn, lower incremental beliefs scores about male-favouring abilities than male non-STEM students ($M = 4.84$, SD = 1.06), $t(122) = 2.47$, $p = 0.008$, $d = 0.46$. Moreover, female STEM students believed more to improve in female-favouring abilities ($M = 4.38$, SD = 0.96) than male STEM students ($M = 4.05$, SD = 1.10), $t(130) = 1.81$, $p = 0.036$, $d = 0.32$, and more than female non-STEM students ($M = 4.08$, SD = 0.92), $t(136) = 1.87$, $p = 0.032$, $d = 0.32$, which did not differ from male non-STEM students ($M = 4.35$, SD = 0.94), see Fig. 1c, d.

This effect was particularly pronounced in the Norwegian sample, as evidenced by a significant three-way interaction, $F(2, 244) = 4.53$, $p = 0.012$, $\eta_p^2 = 0.04$ and $F(2, 244) = 3.93$, $p = 0.021$, $\eta_p^2 = 0.03$, for male-favouring and female-favouring abilities, respectively. Norwegian female STEM students were more convinced ($M = 5.24$, SD = 1.09 for male-favouring abilities and $M = 4.76$, SD = 0.99 for female-favouring abilities) than female non-STEM students ($M = 4.04$, SD = 0.84 for male-favouring abilities and $M = 3.90$, SD = 0.89 for female-favouring abilities) that performance in gender-typed cognitive domains is modifiable, $t(39) = 3.93$, $p < 0.001$, $d = 1.23$ and $t(39) = 2.91$, $p = 0.006$, $d = 0.91$) for male-favouring and female-favouring abilities, respectively. Norwegian male non-STEM students ($M = 4.71$, SD = 0.94) were more convinced than male STEM students ($M = 4.00$, SD = 1.10), $t(39) = 2.20$, $p = 0.034$, $d = 0.70$, that female-favouring cognitive domains are modifiable. Finally, Italian students self-reported a lower incremental belief score about male-favouring abilities ($M = 4.28$, SD = 0.98) than the UK ($M = 4.77$, SD = 0.92) and Norwegian samples ($M = 4.83$, SD = 1.11) as shown by a significant main effect country, $F(2, 244) = 8.26$, $p < 0.001$, $\eta_p^2 = 0.06$.

Cognitive performance

As predicted, men ($M = 12.64$, SD = 4.73) outscored women ($M = 8.64$, SD = 4.42) in MRT, while women ($M = 40.20$, SD = 9.74) outscored men ($M = 37.81$, SD = 9.71) in VF, as indicated by two significant main effects of gender, $F(1, 244) = 45.14$, $p < 0.001$, $\eta_p^2 = 0.16$ and $F(1, 244) = 4.50$, $p = 0.035$, $\eta_p^2 = 0.02$, respectively. The male advantage in mental rotation is $d = 0.88$ and the female advantage in VF was $d = 0.25$. Moreover, STEM students ($M = 11.65$, SD = 5.11) outscored non-STEM students ($M = 9.24$, SD = 4.53) in mental rotation, $F(1, 244) = 14.04$, $p < 0.001$, $\eta_p^2 = 0.05$, $d = 0.50$, while no difference between academic degree was observed for VF, $F(1, 244) = 0.08$, $p = 0.774$,

Table 4 Descriptive statistics for mental rotation (MRT) and verbal fluency (VF) scores split by degree and gender

Task		MRT			VF		
Degree	Gender	M	SD	Range	M	SD	Range
Non-STEM	Women	7.66	3.97	1–23	40.33	9.79	17–62
	Men	11.51	4.34	2–21	38.08	9.34	22–63
STEM	Women	9.74	4.67	0–23	40.06	9.75	18–73
	Men	13s.51	4.86	2–23	37.60	10.05	17–61

$\eta_p^2 < 0.01$, see Table 4 for mean values, standard deviations, and ranges.

Additionally, the main effect country was significant for both MRT, $F(2, 244) = 14.99$, $p < 0.001$, $\eta_p^2 = 0.11$, and VF, $F(1, 244) = 27.70$, $p < 0.001$, $\eta_p^2 = 0.19$. Students from the Italian sample ($M = 8.34$, SD = 3.99) scored lower in the MRT than students from both the UK ($M = 11.48$, SD = 5.25, $p < 0.001$) and Norwegian samples ($M = 11.82$, SD = 4.96, $p < 0.001$). In VF, students from the UK sample ($M = 45.19$, SD = 10.32) outperformed students from the Norwegian ($M = 35.96$, SD = 7.63, $p < 0.001$) and Italian samples ($M = 36.27$, SD = 8.34, $p < 0.001$). Finally, there was a significant interaction between academic degree and country for the MRT, $F(2, 244) = 3.24$, $p = 0.041$, $\eta_p^2 = 0.03$, showing that particularly in the UK sample non-STEM students ($M = 9.08$, SD = 4.05) scored lower than STEM students ($M = 13.66$, SD = 5.25), $d = 0.98$, $p < 0.001$. The mental rotation advantage for STEM students in the Italian ($d = 0.24$) and Norwegian sample ($d = 0.32$) was non-significant ($p > 0.05$). None of the other main effects or interactions were significant (all Fs ≤ 1.28, all ps ≥ 0.279).

Relationships between gender stereotypes, incremental beliefs, and cognitive performance

Across all participants, male-favouring gender stereotypes correlated negatively with female-favouring gender stereotypes [$r(256) = -0.289$, $p < 0.001$]. Endorsed male-favouring gender stereotypes correlated negatively with incremental beliefs regarding both male-favouring [$r(256) = -0.232$, $p < 0.001$] and female-favouring abilities [$r(256) = -0.163$, $p = 0.009$]. Moreover, the two incremental beliefs scores

correlated with each other: $r(256) = 0.661$, $p < 0.001$. Endorsed female-favouring gender stereotypes correlated positively with incremental beliefs about male-favouring abilities [$r(256) = 0.241$, $p < 0.001$] and with the MRT score [$r(256) = 0.154$, $p = 0.013$]. Incremental beliefs about male-favouring abilities correlated with the MRT score: $r(256) = 0.131$, $p = 0.036$. Finally, MRT and VF scores correlated positively, $r(256) = 0.207$, $p = 0.001$.

Considering the correlations separately for gender and degree, it emerged that male-favouring gender stereotypes correlated negatively with MRT performance only for female STEM students $r(62) = -0.252$, $p = 0.044$, indicating that the less pronounced the gender stereotypes, the higher the performance. Moreover, gender stereotypes and incremental beliefs related negatively, but only for male students: STEM [$r(64) = -0.331$, $p = 0.007$], and non-STEM [$r(48) = -0.342$, $p = 0.015$].

The same correlations run separately by country revealed a few significant effects which are detailed in Table 5. However, since there are 24 small subgroups, these correlations are susceptible to multiple testing and outliers. In fact, the only correlation that survived a more conservative alpha criterion of $p = 0.01$ (i.e., the correlation between MRT scores and endorsed male-favouring stereotypes in Italian STEM students) is affected by one outlier. If the outlier is removed, the correlation did not survive alpha-correction for multiple testing, $r(25) = -0.393$, $p = 0.043$.

Table 5 Pearson correlations between gender stereotypes favouring males and mental rotation (MRT)/verbal fluency (VF) performance

Country	Italy		UK		Norway	
Task	MRT	VF	MRT	VF	MRT	VF
Arts students						
Women	−0.514**	0.149	0.251	0.157	0.035	0.072
Men	0.462	0.123	−0.325	<0.001	−0.049	0.434
Science students						
Women	−0.094	0.140	−0.462*	−0.172	−0.188	0.081
Men	0.033	0.151	−0.093	−0.468*	−0.124	0.213

P values are uncorrected. **$p < 0.01$, *$p < 0.05$

Discussion

This study investigated gender stereotypes, incremental beliefs, and cognitive performance in two gender-sensitive tasks, as well as the relationships between these constructs in male and female STEM and non-STEM students in three Western European countries that differ in the gender gap index. The two main propositions were that (a) gender stereotypes and incremental beliefs differ across male and female STEM and non-STEM students and countries, and (b) gender stereotypes and incremental beliefs are associated with cognitive performance in gender-sensitive tasks. Below, we will discuss each of them on the basis of the main results; that is, (a) men endorse male-favouring stereotypes more than women, and women endorse female-favouring stereotypes more than men in the country with the large gender gap, (b) male STEM students endorse male-favouring stereotypes more than female and male non-STEM students, and male non-STEM students endorse female-favouring stereotypes less than female non-STEM and male STEM students, (c) female STEM students believe more in the modifiability of stereotypical male-favouring or female-favouring abilities than female non-STEM students, (d) men outscore women in mental rotation, and women outscore men in verbal fluency, (e) STEM students outperform non-STEM students in mental rotation, and (f) female STEM students' MRT performance increases as male-favouring stereotypes decrease.

Stereotypes favouring one's own gender are endorsed more

Overall, female students reported to endorse female-favouring stereotypes more than male students, while male students endorsed male-favouring stereotypes more than female students. In other words, each gender endorsed more strongly stereotypes that highlight one's own superior skills. The effect size was in the 'small' range, that is 0.34 and 0.48 for male-favouring and female-favouring abilities, respectively.

A significant interaction with country showed that this applied mostly for the country with the large gender gap, leading to a medium-effect size ($d = 0.75$) for the female-favouring stereotypes and a large-effect size ($d = 0.91$) for the male-favouring stereotypes. This suggests that living in a country with a large gender gap might be associated with stronger gender stereotypes and, possibly, such a tendency toward stereotyping could contribute to sustaining the gender gap.

Male STEM students are more stereotyped

First, it should be pointed out that all groups, male and female, STEM and non-STEM students, endorsed gender stereotypes with respect to stereotypical male-favouring or female-favouring abilities as shown by the significant intercept. However, the magnitude varied: male STEM students endorsed male-favouring stereotypes more strongly than female STEM students (a medium-effect size of $d = 0.66$) and male non-STEM students (a small-effect size of $d = 0.45$). Male non-STEM students endorsed female-favouring stereotypes less than male STEM students (a small-effect size of $d = 0.47$) and female non-STEM students (a large-effect size of $d = 0.88$). In the present study, male-favouring and female-favouring abilities were tested with two typical measures: mental rotation and verbal fluency. Men outperformed women in mental rotation, and women outperformed men in verbal fluency. Moreover, male STEM students performed better than female STEM students and any non-STEM group. Thus, the stronger endorsement of male-favouring abilities in male STEM students and of female-favouring abilities in female students aligns with their better mental rotation and verbal fluency skills, respectively. Gender stereotypes need not be problematic per se as long as they are recognized as such: overgeneralizations and oversimplifications that may be grounded in reality, in part, and that might be informative about group differences on average in certain tasks. However, these gender stereotypes can become problematic, if they lead to the erroneous expectation that all women perform poorly in male-favouring subjects, such as mathematics and spatial abilities, or that all men perform poorly in female-favouring subjects. This would, indeed, become fertile ground for a "chilly climate" that prevents women from taking STEM degrees (and men from non-STEM degrees), as has been found recently (e.g., Cabay, Bernstein, Rivers, & Fabert 2018, p. 3). Interventions focused on fostering students' sense of belonging, self-affirmation techniques (reminding a positive self-image), shaping effective role models, and raising self-efficacy can be effective resulting in higher levels of confidence, performance, career advancement, as well as identification and friendship (e.g., Carnes et al. 2015; Walton, Logel, Peach, Spencer, & Zanna 2015). Another strategy could be to make students aware (e.g., by pointing out scientific evidence) that facing and exercising with spatial school subjects (Moè 2016b) and experience with spatial activities favour performance in male-typical domains (Newcombe & Frick 2010). In turn, exercising with verbal tasks is likely to improve verbal abilities.

Female STEM students have higher incremental beliefs

Means in incremental beliefs ranged between 4.05 and 4.84 on a scale from 1 to 7, implying that on average, participants held similar incremental beliefs. However, we found a significant interaction between gender and academic degree: Female STEM students believed more in the modifiability

of stereotypical female-favouring abilities than female non-STEM and male STEM students (ds = 0.32). In turn, female non-STEM students held lower incremental beliefs than male and female STEM students ($d = 0.46$ and $d = 0.33$, respectively). The interaction patterns in panels c and d in Fig. 1 suggest that studying a subject, which is traditionally perceived to be the domain of the opposite gender (e.g., women pursuing an STEM degree and men a non-STEM degree) is associated with increased beliefs that one can improve in those domains. There are two possibilities for this finding: perhaps, women enrolled in STEM degrees and men enrolled in non-STEM degrees adjust their gender stereotypes and incremental beliefs as a consequence of more successful practice with stereotypical male-tasks. Alternatively, those men and women might have already higher incremental beliefs before their enrolment and, thus, were more confident to enter a degree that is stereotypically perceived to be the domain of the opposite gender.

Cognitive performance differs between genders and degrees

As expected, men outperformed women in mental rotation, and women outperformed men in verbal fluency. The magnitude of the gender differences in mental rotation ($d = 0.88$) and verbal fluency ($d = 0.25$) was perfectly in line with the literature (Linn, & Petersen 1985; Voyer et al. 1995; Hyde & Linn 1988). Interestingly, the male advantage in MRT and the female advantage in verbal fluency were independent of academic degree, implying that men in both STEM and non-STEM degrees outperformed their female counterparts and that women in both STEM and non-STEM degrees outperformed their male counterparts. This contradicts the hypothesis that female STEM students and male non-STEM students have similar mental rotation and verbal fluency skills as their respective male STEM and female non-STEM peers. Furthermore, this speaks to the robustness of the male and female advantage in mental rotation and verbal fluency, respectively, whose developmental origins lie probably well before individuals choose an academic degree.

Moreover, we also corroborated the previous findings that STEM students obtain higher MRT scores than non-STEM students. Only a few studies (e.g., Hausmann 2014; Sanchis-Segura et al. 2018) compared STEM vs. non-STEM students' mental rotation performance, while in other studies, degree was not specified (e.g., Halari et al. 2005), or only students from one single discipline (mainly psychology) were recruited (e.g., Moè et al. 2009; Moè 2018b). We did not replicate the advantage of non-STEM students in verbal fluency reported by Hausmann (2014). This may be due to sample characteristics, such as size and admission criteria in the three involved universities. Further studies are needed

to clarify whether non-STEM students are more proficient in verbal skills.

Gender stereotypes and incremental beliefs relate weakly with performance

We found that the higher the mental rotation score, the higher the belief in modifiability of male-favouring abilities. Moreover, it increased in female STEM students the lower their endorsement of male-favouring gender stereotypes. This is in accordance with the previous results, suggesting that priming gender stereotypes affects performance in cognitive tests (e.g., Moè, 2009; Hirnstein et al., 2012), and that the higher the beliefs in modifiability, the higher the mental rotation score (Moè et al., 2009).

Moreover, this suggests that women studying STEM degrees tend to adjust their gender stereotypes in accordance with cognitive performance: if they perform well, they might believe less in stereotypes, but if they do not perform well, they appear to externalize, attributing their poor performance to gender stereotypes.

Unexpectedly, the typical negative relationship between gender stereotypes and incremental beliefs [i.e., the more a person holds a stereotype, the less s(he) believes that abilities can be changed] was found only in men. This shows that incremental beliefs in women were relatively unaffected by gender stereotypes, suggesting that women can develop the belief to improve in male-favouring tasks independently from the stereotypes they hold.

Gender gaps could be linked to endorsed stereotypes and incremental beliefs

Gender stereotypes were endorsed more in the country with the larger gender gap. This confirms the hypothesis that the larger the gender gap, the larger the endorsed gender stereotypes. Again, as support for our hypothesis, students in the country with the larger gender gap had lower incremental beliefs scores, suggesting that they are less convinced that the ability in the cognitive task, in which men typically perform better than women, can be improved. On the other hand, in the country with the smallest gender gap, female STEM students believed more than non-STEM students that male-favouring and female-favouring abilities can be changed. This result suggests that the factor gender gap might play a role mainly in conjunction with the degree chosen and/or familiarization with cognitive tasks that are mostly recognized as male cognitive domains. Finally, it is important to note that the general difference between countries in cognitive performances might well be due to different admission criteria at the three universities rather than related to the gender gap index.

To the best of our knowledge, we are not aware of any cross-cultural studies that examined endorsed gender stereotypes, incremental beliefs, and cognitive performances within the same study, and this is the first attempt to investigate differences among three countries with respect to the gender gap index. The results of the present study supported our hypothesis that a larger gender gap might increase cognitive gender differences, at least as far as endorsed stereotypes about male-favouring abilities and incremental beliefs are concerned.

Limitations and future directions

First, one must interpret the findings with caution regarding the female-favouring stereotype score and the incremental belief score about female-favouring abilities due to the low internal consistency of the female-favouring abilities in the Beliefs questionnaire, which is probably due to the wide range of tasks included. Future research could consider a more homogeneous set of abilities, perhaps including only verbal tasks (i.e., verbal fluency, anagrams, finding a rhyme, etc.). Second, the differences between countries do not necessarily allow to attribute those findings to differences in gender equality. The current study tested participants from only one university in each country. Therefore, generalizing the effects to the entire country is not justified. For example, it is conceivable that different admission criteria for the three universities explain the differences in cognitive performance better than country-specific differences in gender stereotypes and incremental beliefs. In the present study, we recruited participants from only one university in each of the three different countries as part of an existing collaboration and with the aim to increase the generalizability of our findings. However, we hope that the current study will encourage researchers to replicate findings with larger samples and at other universities/countries. Third, this is a cross-sectional study comparing students who are enrolled in different subjects. Longitudinal studies have hardly been conducted in this area (Wang & Degol 2013), with a few exceptions regarding ability perception (Simpkins, Davis-Kean, & Eccles 2006), and implicit theories (Blackwell, Trzesniewski, & Dweck 2007). Thus, it is unclear to what extent students changed their gender stereotypes and incremental beliefs after entering an STEM degree. Similarly, it is unclear to what extent students with particularly low gender stereotypes and high incremental beliefs took up an STEM degree, as suggested by Leslie et al. (2015). Future longitudinal studies need to address this issue.

Conclusions

The findings of the present study revealed that gender stereotypes with respect to stereotypical male-favouring and female-favouring abilities are prevalent among men and women, respectively, especially in the country with the largest gender gap. Moreover, male STEM students were found to endorse more male-favouring stereotypes and male non-STEM students to endorse less female-favouring stereotypes in comparison with female students. Additionally, female STEM students believed more in the modifiability of typical female-favouring abilities, while male non-STEM students were more convinced than females that female-favouring abilities can be improved. Gender stereotypes and incremental beliefs were only modestly related to cognitive performance in tasks known to yield reliable male and female advantages.

In sum, our results showed that male STEM students endorse stronger male-favouring gender stereotypes which are at least partly in line with the better mental rotation skills in men than women. This could lead to the erroneous conclusion that women in general lack the capacity to study and thrive in STEM subjects. However, the stronger conviction in female STEM students, as compared to female non-STEM students, that male-favouring abilities are modifiable, demonstrates their belief that they can improve and succeed in typical male-favouring tasks.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical statements All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committees and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Andreano, J. M., & Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learning & Memory, 16*, 248–266. https://doi.org/10.1101/lm.918309.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x.

Bull, R., & Benson, P. J. (2006). Digit ratio (2D:4D) and the spatial representation of magnitude. *Hormones and Behavior, 50*, 194–199. https://doi.org/10.1016/j.yhbeh.2006.02.008.

Bull, R., Cleland, A. A., & Mitchell, T. (2013). Sex differences in the spatial representation of number. *Journal of Experimental Psychology: General, 142*, 181–192. https://doi.org/10.1037/a0028387.

Cabay, M., Bernstein, B. L., Rivers, M., & Fabert, N. (2018). Chilly climates, balancing acts, and shifting pathways: What happens to women in STEM doctoral programs. *Social Sciences, 7*, 23. https://doi.org/10.3390/socsci7020023.

Carnes, M., Devine, P. G., Baier Manwell, L., Byars-Winston, A., Fine, E., Ford, C. E., et al. (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*, 221–230. https://doi.org/10.1097/ACM.0000000000000552.

Ceci, S. J. (2017). Women in academic science: Experimental findings from hiring studies. *Educational Psychologist, 53*, 22–41. https://doi.org/10.1080/00461520.2017.1396462.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest, 15*, 75–141. https://doi.org/10.1177/1529100614541236.

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

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math—Gender stereotypes in elementary school children. *Child Development, 82*, 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x.

Dweck, C. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Eddy, S. L., & Brownell, S. E. (2016). Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Physical Review Physics Education Research, 12*, 1–20. https://doi.org/10.1103/PhysRevPhysEducRes.12.020106.

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*, 175–191.

Fredericks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*, 519–533. https://doi.org/10.1037/0012-1649.38.4.519.

Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., et al. (2005). Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins. *Behavioral Neuroscience, 119*, 104–117. https://doi.org/10.1037/0735-7044.119.1.104.

Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed.). NY: Psychology Press.

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition, 45*, 392–414. https://doi.org/10.1006/brcg.2001.1287.

Hausmann, M. (2014). Arts versus science—Academic background implicitly activates gender stereotypes on cognitive abilities with threat raising men's (but lowering women's) performance. *Intelligence, 46*, 235–245. https://doi.org/10.1016/j.intell.2014.07.004.

Hausmann, M., Schoofs, D., Rosenthal, H. E., & Jordan, K. (2009). Interactive effects of sex hormones and gender stereotypes on cognitive sex differences: A psychobiosocial approach. *Psychoneuroendocrinology, 34*, 389–401. https://doi.org/10.1016/j.psyneuen.2008.09.019.

Heil, M., Jansen, P., Quaiser-Pohl, C., & Neuburger, S. (2012). Gender-specific effects of artificially induced gender beliefs in mental rotation. *Learning and Individual Differences, 22*, 350–353. https://doi.org/10.1016/j.lindif.2012.01.004.

Herlitz, A., Airaksinen, E., & Nordström, E. (1999). Sex differences in episodic memory: The impact of verbal and visuospatial ability. *Neuropsychology, 13*, 590–597. https://doi.org/10.1037//0894-4105.13.4.590.

Hirnstein, M., Coloma Andrews, L. C., & Hausmann, M. (2014). Gender-stereotyping and cognitive sex differences in mixed and same-sex groups. *Archives of Sexual Behavior, 43*, 1663–1673. https://doi.org/10.1007/s10508-014-0311-5.

Hirnstein, M., Freund, N., & Hausmann, M. (2012). Gender stereotyping enhances verbal fluency performance in men (and women). *Zeitschrift für Psychologie, 220*, 70–77. https://doi.org/10.1027/2151-2604/a000098.

Horn, W. (1962). *Leistungsprüfsystem (LPS)*. Göttingen: Hogrefe Verlag fuer Psychologie.

Huber, S., Nuerk, H. C., Reips, U. D., & Soltanlou, M. (2017). Individual differences influence two-digit number processing, but not their analog magnitude processing: A large-scale online study. *Psychological Research Psychologische Forschung, 83*, 1444–1464. https://doi.org/10.1007/s00426-017-0964-5.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal abilities: A meta-analysis. *Psychological Bulletin, 104*, 53–69.

Kurtz-Costes, B., Copping, K. E., Rowley, S. J., & Kinlaw, C. R. (2014). Gender and age differences in awareness and endorsement of gender stereotypes about academic abilities. *European Journal of Psychology of Education, 29*, 603–618. https://doi.org/10.1007/s10212-014-0216-7.

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

Levine, S. C., Foley, A., Lourenco, S., Ehrich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Wiley Interdisciplinary Review Cognitive Science, 7*, 127–155. https://doi.org/10.1002/wcs.1380.

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*, 1421–1436. https://doi.org/10.1037/0022-3514.74.6.1421.

Lezak, M., Howieson, D., Bigler, E., & Tranel, D. (2012). *Neuropsychological Assessment*. New York, NY: Oxford University Press.

Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research, 41*, 63–76. https://doi.org/10.1080/0013188990410106.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex-differences in spatial ability—A meta-analysis. *Child Development, 56*, 1479–1498.

Lugli, L., D'Ascenzo, S., Borghi, A. M., & Nicoletti, R. (2018). Clock walking and gender: How circular movements influence arithmetic calculations. *Frontiers in Psychology, 9*, 1599. https://doi.org/10.3389/fpsyg.2018.01599.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18*, 37–45. https://doi.org/10.1016/j.tics.2013.10.011.

Moè, A. (2009). Expectations and recall of texts: The more able–more difficult effect. *Learning and Individual Differences, 19*, 609–614. https://doi.org/10.1016/j.lindif.2009.08.002.

Moè, A. (2012). Gender difference does not mean genetic difference: Externalizing improves performance in mental rotation. *Learning and Individual Differences, 22*, 20–24. https://doi.org/10.1016/j.lindif.2011.11.001.

Moè, A. (2016a). Teaching motivation and strategies to improve mental rotation abilities. *Intelligence, 59*, 16–23. https://doi.org/10.1016/j.intell.2016.10.004.

Moè, A. (2016b). Does experience with spatial school subjects favour girls' mental rotation performance? *Learning and Individual Differences, 47*, 11–16. https://doi.org/10.1016/j.lindif.2015.12.007.

Moè, A. (2018a). Mental rotation and mathematics: Gender-stereotyped beliefs and relationships in primary school children. *Learning and Individual Differences*, *61*, 172–180. https://doi.org/10.1016/j.lindif.2017.12.002.

Moè, A. (2018b). Effects of group gender composition on Mental Rotation Test performance in women. *Archives of Sexual Behavior*, *47*, 2299–2305. https://doi.org/10.1007/s10508-018-1245-0.

Moè, A., Jansen, P., & Pietsch, S. (2018). Childhood preference for spatial toys. Gender differences and relationships with mental rotation in STEM and non-STEM students. *Learning and Individual Differences, 68*, 108–115. https://doi.org/10.1016/j.lindif.2018.10.003.

Moè, A., Meneghetti, C., & Cadinu, M. (2009). Women and mental rotation: Incremental theory and spatial strategy use enhance performance. *Personality and Individual Differences, 46*, 187–191. https://doi.org/10.1016/j.paid.2008.09.030.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*, 369–377. https://doi.org/10.1016/j.lindif.2007.01.002.

National Science Foundation. (2018). Figure 3–26. *Women in the workforce and in S&E: 1993 and 2015*. Retrieved from www.nsf.gov/statistics/2018/nsb20181/report/sections/science-and-engineering-labor-force/women-and-minorities-in-the-s-e-workforce.

Neill, J. (2008). *Writing up a factor analysis*. Retrieved on October 24th www.bwgriffin.com/gsu/courses/edur9131/content/Neill2008_WritingUpAFactorAnalysis.pdf.

Newcombe, N. S., & Frick, A. (2010). Early education for spatial intelligence: Why, what, and how. *Mind, Brain, and Education, 4*, 102–111.

North, S. (2005). Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Studies in Higher Education, 30*, 517–533. https://doi.org/10.1080/03075070500249153.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math _ male, me _ female, therefore math not _ me. *Journal of Personality and Social Psychology, 83*, 44–59. https://doi.org/10.1037/0022-3514.83.1.44.

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal, 48*, 1125–1156. https://doi.org/10.3102/0002831211410683.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39–58.

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four cross-cultural samples (N = 3367): Overall sex differences and the role of academic program in performance. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 42*, 1005–1014. https://doi.org/10.1016/S0010-9452(08)70206-5.

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS ONE, 7*, e39904. https://doi.org/10.1371/journal.pone.0039904.

Sanchis-Segura, C., Aguirre, N., Cruz-Gómez, Á. J., Solozano, N., & Forn, C. (2018). Do gender-related stereotypes affect spatial performance? Exploring when, how and to whom using a chronometric two-choice Mental Rotation Task. *Frontiers in Psychology, 9*, 1261–1261. https://doi.org/10.3389/fpsyg.2018.01261.

Scheuringer, A., Wittig, R., & Pletzer, B. (2017). Sex differences in verbal fluency: The role of strategies and instructions. *Cognitive Processing, 18*, 407–417. https://doi.org/10.1007/s10339-017-0801-1.

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

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703. https://doi.org/10.1126/science.171.3972.701.

Simpkins, S. D., Davis-Kean, P., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42*, 70–83. https://doi.org/10.1037/0012-1649.42.1.70.

Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles, 67*, 617–629. https://doi.org/10.1007/s11199-012-0209-4.

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS ONE, 8*, 10. https://doi.org/10.1371/journal.pone.0057988.

Szameitat, A. J., Hamaida, Y., Tulley, R. S., Saylik, R., & Otermans, P. C. J. (2015). "Women are better than men"—Public beliefs on gender differences and other aspects in multitasking. *PLoS ONE, 10*, e0140371. https://doi.org/10.1371/journal.pone.0140371.

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental psychology, 47*, 943–949. https://doi.org/10.1037/a0024047.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604.

Vander Heyden, K. M., van Atteveldt, N. M., Huizinga, M., & Jolles, J. (2016). Implicit and explicit gender beliefs in spatial ability: Stronger stereotyping in boys than girls. *Frontiers in Psychology, 7*, 1114–1114. https://doi.org/10.3389/fpsyg.2016.01114.

Voyer, D., & Jansen, P. (2016). Sex differences in chronometric mental rotation with human bodies. *Psychological Research Psychologische Forschung, 80*, 974–984. https://doi.org/10.1007/s00426-015-0701-x.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250–270.

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology, 107*, 468–485. https://doi.org/10.1037/a0037461.

Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33*, 1–49. https://doi.org/10.1016/j.dr.2013.08.001.

Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and

mathematics. *Psychological Science, 24*, 770–775. https://doi.org/10.1177/0956797612458937.

Watkins, M. W. (2000). *Monte Carlo PCA for Parallel Analysis* [computer software]. State College, PA: Ed & Psych Associates.

World Economic Forum. (2016). *The Global Gender Gap Report*. Retrieved from November 19th 2019 www.weforum.org/reports/the-global-gender-gap-report-2016.

World Economic Forum. (2018). *The Global Gender Gap Report*. Retrieved from November 19th 2019 https://reports.weforum.org/global-gender-gap-report-2018.

Wraga, M., Duncan, L., Jacobs, E. H., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13*, 813–819. https://doi.org/10.3758/BF03194002.

Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist, 70*, 10–20. https://doi.org/10.1037/a0038208.