



# Spatial Skills and Self-Perceived Masculinity: Considering College Major STEM-ness, Directionality, and Gender

Dominic P. Kelly<sup>1</sup> · Adriene M. Beltz<sup>1</sup>

Accepted: 3 June 2022 / Published online: 10 September 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

The sex-role mediation hypothesis suggests that a masculine self-concept promotes male-typed cognition, including spatial skills. Support for the hypothesis is mixed, limited by small samples and the spatial skills examined, with few studies exploring the role of gendered activities, experiences, and interests (e.g., Science, Technology, Engineering, and Mathematics [STEM] college majors). Therefore, in a sample of 339 undergraduate students, a series of regression analyses with bootstrapped-based estimation of indirect effects was used to determine whether self-perceived masculinity was related to three-dimensional (3D) mental rotations, geographical knowledge, identifying the true horizontal, and object location memory via major ‘STEM-ness.’ Spatial skills and masculinity were consistently positively related, except for object location memory, which is the only spatial skill examined where women, on average, outperform men. Moreover, the link between some spatial skills (3D mental rotations, identifying the true horizontal) and masculinity partially occurred via major STEM-ness. Findings are novel in revealing associations among masculinity, spatial skills, and STEM interests, and are somewhat consistent with the sex-role mediation hypothesis. They also encourage future longitudinal studies to examine whether masculinity predicts or is predicted by spatial skills, and they may have downstream implications for reducing gender disparities in STEM.

**Keywords** Academic specialization · Gender expression · Masculinity · Mental rotation · Spatial ability

Gender is multidimensional and reflects a wide range of separate but related constructs (Ruble et al., 2006). Individuals are unlikely to be equally gender-typed in all constructs; for example, a girl may be highly gender-typed in her beliefs (e.g., thinking math skills are more important for boys than for girls) but not in her behaviors (e.g., being highly physically aggressive). Also, individuals of the same gender are not equally gender-typed as there is substantial variation among girls that overlaps with variation among boys (Hyde, 2005). Thus, studying gender constructs in isolation can lead to both knowledge gaps and inaccuracies. Although there are some theories about how constructs are related, empirical investigation is relatively limited.

An important example concerns the relation between gender self-concept and gendered cognition. Gender self-concept concerns how individuals perceive gender, as influenced by both their individual feelings and sociocultural

experiences, and is apparent in the ways that individuals use gender labels for themselves (and potentially others), including reports of their own self-perceived masculinity and femininity (Ruble et al., 2006; Wood & Eagly, 2015). Gendered cognition concerns the skills individuals display in domains that typically show a gender difference. For example, there are no gender differences in overall intelligence, but – on average – men outperform women on spatial tasks (Voyer, 2011) and women outperform men on language tasks (Halpern, 2013). These patterns are not absolute, though, as the qualitative pattern of gender differences varies across spatial skills: average differences favor women in some spatial tasks, such as object location memory, and favor men in some language tasks, such as verbal analogies; thus, both spatial and language skills are clearly multidetermined (Beltz et al., 2020; Halpern, 2013; Voyer et al., 2007). The sex-role mediation hypothesis (Nash, 1979) suggests that – regardless of sex or gender – gender self-concept and cognition are yoked, such that masculinity facilitates performance in spatial tasks, and that femininity facilitates performance in language tasks.

✉ Adriene M. Beltz  
abeltz@umich.edu

<sup>1</sup> Department of Psychology, University of Michigan,  
Ann Arbor, MI, USA

There is some empirical support for the hypothesis. For instance, an early meta-analysis of 338 effect sizes from various spatial and mathematics tasks revealed an overall positive relation with masculinity (Signorella & Jamison, 1986). More recently, another meta-analysis (including studies since 1986) also revealed a positive association between masculinity and mental rotations separately in both men ( $r=0.30$ ) and women ( $r=0.23$ ) (Reilly & Neumann, 2013). Additionally, a follow-up empirical study with a large sample size ( $N=309$ ) and several different spatial skills (i.e., two-dimensional [2D] and three-dimensional [3D] mental rotations tasks, the Piagetian Water Level task, and a Group Embedded Figures Task) reported that masculinity was positively correlated with all spatial skills except for 2D mental rotations – including the Group Embedded Figures Task, a task that typically does not show a gender difference (Reilly et al., 2016).

Despite the intuitive appeal of the sex-role mediation hypothesis and the presumptive empirical evidence, the supporting data have several limitations. First, mental rotations is studied most frequently (with enough reports for the meta-analysis described above; Reilly et al., 2016), so less is known about other spatial skills, including those that show a mean-level difference favouring women (e.g., object location memory; Voyer et al., 2007). Second, mechanisms underlying the relation between gender self-concept and gendered cognition are rarely empirically tested, but are important to reveal to inform potential interventions. Third, the extant research assumes that gender self-concept *leads to* gendered cognition and fails to consider the opposite direction of effects. Might gendered cognition influence gender self-concept? Fourth, gender differences in the relation between gendered self-concept and cognition are inconsistently investigated (and found), so it is unclear whether and how the link between these gendered constructs has implications for reducing gender disparities in spatial and language tasks. Thus, the aims of this largest-to-date study are to: (1) examine the relation between masculinity and a variety of spatial skills; (2) test a potential mechanism underlying the relation, namely STEM-ness of college major, which reflects gendered activities, experiences, and interests; (3) explore the direction of the relation, and (4) explore whether the relation differs in degree or form between young women and men.

## Knowledge Gaps in the Sex-Role Mediation Hypothesis

Most studies on the sex-role hypothesis and spatial skills have indexed gendered cognition solely by mental rotations. Mental rotations tasks require comparisons between a 2D or 3D target shape and an array of similar shapes; the goal is to identify which shape(s) in the array are rotated versions

of the target. This is evidenced by the most recent meta-analysis on the topic, which included multiple measures of gender self-concept (e.g., Bem Sex Role Inventory, Personal Attributes Questionnaire; Bem, 1974; Spence et al., 1974), but only mental rotations as a measure of gendered cognition (Reilly & Neumann, 2013). Although the focus on mental rotations is reasonable because it shows the largest gender difference in spatial cognition (Beltz et al., 2020) and is a widely used task to study spatial cognition, it nonetheless begs the question of whether findings generalize to other spatial skills, including those that at a mean level favor women (e.g., object location memory; Voyer et al., 2007).

In addition, mechanisms underlying the sex-role hypothesis and spatial skills are unclear. Some have speculated that the gendered nature of activities matters (Reilly & Neumann, 2013). Masculine play in childhood is generally space-occupying and has been thought to involve manipulation and rotation of objects (Benenson et al., 2011; Newcombe et al., 1983); thus, those who identify or perceive themselves as masculine and engage in masculine activities may indirectly hone their spatial skills. A similar logic applies to feminine identification, feminine activities, and language skills; for instance, feminine play in childhood is thought to rely more on extended conversation (reviewed in Riley & Jones, 2007); therefore, those who identify or perceive themselves as feminine and engage in feminine activities may indirectly hone their language skills. Additionally, given that participation in science, technology, engineering, and mathematics (STEM) fields has a male-biased stereotype and that there is substantial evidence for a positive link between spatial skills and participation in STEM fields (Andersen, 2014; Lubinski, 2010; Wai et al., 2009), STEM participation may similarly indirectly affect the relation between perceived masculinity and spatial skills. This mechanism is relatively unexplored in the literature and requires empirical testing.

Furthermore, consistent with the original sex-role mediation hypothesis, most research assumes that gender self-concept *leads to* gender-typed cognition, but this may be a critical oversight. The basic notion of the hypothesis is that those who identify as boys/men engage in masculine activities, indirectly honing their spatial skills. Alternatively, given the endorsed stereotype that men are better at spatial tasks (Halpern et al., 2011), experiences excelling at spatial tasks could reinforce a masculine self-concept and negative experiences with spatial tasks could challenge a masculine self-concept (McGlone & Aronson, 2006). Thus, gendered cognition might influence gender self-concept. It is also possible that effects are bidirectional: "... it remains possible that competencies for intellectual tasks help further refine one's sex-role identity, or that there are bidirectional links between sex-role identity and intellectual abilities" (Reilly et al., 2016, p. 156). This is consistent with the dual-pathways model of gender differentiation, or that children are both affected by

gendered beliefs that affect their experiences and shape their own beliefs based on their experiences (Liben & Bigler, 2002). Unfortunately, the extant literature only examines the relation from self-concept to cognition, leaving this question wholly unanswered.

Lastly, there is reason to suspect that the sex-role mediation hypothesis may not unfold to the same degree or through the same processes in boys/men and girls/women. For instance, in the first major meta-analysis on the topic, effects of masculinity were stronger for women than men, and there was no overall positive relation between femininity and language task performance in another 72 effect sizes (Signorella & Jamison, 1986). The former gender difference was not observed in the second major meta-analysis on the topic (Reilly & Neumann, 2013). The role of gender in the sex-role mediation hypothesis is difficult to decipher from recent empirical studies, though, as gender differences are assumed to be only *quantitative* (i.e., reflecting the same process that differs in degree or magnitude between men and women); this is evidenced when gender is statistically controlled in analyses (see Beltz et al., 2019). Gender differences could also be inconsistently detected in past work because they are *qualitative* (i.e., reflecting different processes in men and women; Becker & Koob, 2016; Beltz et al., 2019). For instance, based on average patterns in gender differences, the link between masculinity and spatial skills is gender-congruent for men, but gender-incongruent for women. Moreover, there is evidence that men and women use different strategies to complete spatial tasks, as men are more likely than women to use global, holistic processing whereas women are more likely than men to use local processing, to compare specific features (Boone & Hegarty, 2017; Hegarty, 2018). In these cases, mere quantitative comparisons between men and women could obfuscate or cancel out potential differences.

## Current Study

The overarching goal of the present study is to elucidate the relation between gender self-concept and gendered cognition via four aims. The first aim is to examine the consistency of the association between self-perceived masculinity and a variety of spatial skills, including tests of: (a) 3D mental rotations; (b) geographical knowledge requiring identifying locations on a map; (c) spatial perception or the ability to identify the true horizontal; and (d) object location memory or short-term spatial memory of arrays of objects. There is an average gender difference favoring boys/men on the first three skills, but an average difference favoring girls/women in object location memory (see Beltz et al., 2020); thus, patterns of findings will help determine whether the potential relation between masculinity and spatial skills only applies to stereotypically masculine aspects of gender. The second

aim is to examine whether college major ‘STEM-ness’ (reflecting gendered activities, experiences, and interests) partially explains the relation between gender self-concept and gendered cognition. The third aim is to explore whether gender self-concept (indexed by self-perceived masculinity) is a better predictor of gendered cognition (indexed by a variety of spatial skills) or vice versa. Although data are cross-sectional, some statistical insight can be achieved by examining whether more variance is explained by models with self-concept or with spatial skills as outcomes. The fourth aim is to determine whether there is evidence for quantitative and/or qualitative gender differences in the previous aims, which may help elucidate the mechanisms underlying potential gender disparities in spatial skills and may, in turn, inform downstream interventions. These aims will be addressed in the largest dataset on the topic to-date; the average sample size of  $N = 136$  in studies included in a recent meta-analysis (Reilly & Neumann, 2013) is too small to detect the small-to-medium sized effects that are often reported (Cohen, 1992) and raises questions about replication (e.g., Maxwell et al., 2015; Pashler & Harris, 2012).

## Method

Data came from a project concerning gendered behaviour and cognition. A previous report using a subset of these data showed a gender difference in 3D mental rotations performance and revealed exogenous hormone influences on that skill (e.g., for oral contraceptives with specific pharmacokinetic formulations; Beltz et al., 2015).

## Participants and Procedure

Participants were undergraduate students recruited from a subject pool at a large public university in the United States. All participants were at least 18 years of age. There were no inclusion criteria, but some women were recruited based on menstrual cycle regularity or oral contraceptive use. This study was approved by the Institutional Review Boards of the Pennsylvania State University and the University of Michigan. All the procedures in this study are performed in line with the principles of the Declaration of Helsinki.

There were 411 (274 women) participants who completed a 60-min monitored online survey containing a series of questionnaires and computerized cognitive tests in a research laboratory. They received course credit for their participation. In total, 72 (49 women) participants were excluded for the following reasons. First, 55 (43 women) participants were excluded for not having declared a college major (i.e., missing major STEM-ness) which precluded their inclusion in analyses of indirect effects. Second, five participants (three women) were excluded for testing issues, including

distraction or low effort detected via research assistant reports recorded during monitored testing. Third, six participants (three women) were excluded for having scores on the measure of general cognitive ability (described below) below zero, indicating notable problems with comprehension or low effort. Fourth, six men were outliers on age (i.e., over 3 standard deviations above the sample mean). Excluded participants did not significantly differ (i.e.,  $p$ 's > .05) from the remaining sample on age, self-perceived masculinity, major STEM-ness, or spatial skills; however, they did differ on general cognitive ability because six participants were excluded for having negative scores on that test. In addition, missing values analyses were conducted using the SPSS Missing Values add-on module and suggest that the data appear to be missing completely at random, except for major STEM-ness on which younger participants were more likely to be excluded because they had not yet declared a major. Analyses did not suggest any clear patterns of missingness in study variables showed in Table 1 according to participants' race, ethnicity and gender.

The final sample therefore consists of 339 participants (225 women) between 18 and 23 years old ( $M = 19.23$ ,  $SD = 0.96$ ). They self-identified as White (85.8%), Black (4.7%), Asian (4.4%), Native American (0.3%), or multiple races (3.2%), and 0.5% did not respond; 6.8% also identified as Latinx (with 91.4% identifying as non-Latinx and 1.8% not responding). Most were first year students (68.1%), with 20.4% second year, 7.1% third year, 3.2% fourth year and 1.2% fifth year students participating. There was a significant age difference between men and women,  $t(337) = 4.94$ ,  $p < .001$ ,  $d = 0.55$ . Men ( $M = 19.58$ ,  $SD = 1.04$ ) were on average 6 months older than women ( $M = 19.05$ ,  $SD = 0.87$ ).

## Measures

This study concerns self-perceived masculinity (a measure of gender self-concept), four different spatial skills (measures of gendered cognition) that have been widely used in

psychological studies on gender similarities and differences (e.g., Berenbaum et al., 2012; Blakemore et al., 2009), and college major STEM-ness (a proxy for gendered activities, experiences, and interests). General cognitive ability and age were considered as covariates due to their known positive relations with spatial skills and masculinity, respectively (Barrett & Raskin White, 2002; Johnson & Bouchard Jr, 2005).

## Self-Perceived Masculinity

Self-perceived masculinity was assessed with the six-item Sex Role Identity Scale (Storms, 1979), which is a widely used measure of gender self-concept that reflects gender self-categorizations and expression (e.g., Lippa, 2002; Martin & Finn, 2010; Steele et al., 2019). The measure shows convergent validity with other widely used measures of masculinity, including the Minnesota Multiphasic Personality Inventory-2 (Butcher et al., 1989; Johnson et al., 1996), various forms of the Personality Attributes Questionnaire (Di Dio et al., 1996; Hungerford & Sobolew-Shubin, 1987; Spence & Helmreich, 1978; Spence et al., 1979; Storms, 1979), and the Bem Sex-Role Inventory (Bem, 1974; Hungerford & Sobolew-Shubin, 1987; O'Heron & Orlofsky, 1990), among other gender expression measures (e.g., Lehavot et al., 2011). Further, the measure shows the expected correlations with other gender-related constructs, including gender typicality (DiDonato & Berenbaum, 2011), gendered interests and role behaviours (Di Dio et al., 1996) and mental health (O'Heron & Orlofsky, 1990).

Participants were asked to respond to items on a scale between 1 (*not at all*) and 5 (*extremely*) concerning the extent to which they feel masculine and then feminine in general, in their dress and in their actions. Specific example items are, "How masculine do you act, appear, and come across to others?" and "In general, how feminine do you think you are?" Thus, this is a self-report measure of participants' self-defined masculinity and femininity and not

**Table 1** Gender differences in study variables with descriptive statistics

	Women			Men			Gender Difference			
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Masculinity	225	1.81	.74	114	3.67	.59	23.30	337	<.001	2.68
3D mental rotations	216	26.99	6.18	106	31.46	7.41	5.71	320	<.001	.68
Geographical knowledge	220	13.41	2.86	114	14.51	2.30	3.56	332	<.001	.41
Identifying the true horizontal	223	6.48	3.51	113	8.02	3.67	3.74	334	<.001	.43
Object location memory	224	8.76	2.64	112	8.26	2.78	-1.62	334	.053	-.19
Major STEM-ness	225	3.61	.71	114	3.81	.90	2.20	337	.014	.25
General cognitive ability	225	11.05	4.30	114	11.52	4.79	.91	337	.182	.11

All tests were one-tailed due to directional hypotheses, except for general cognitive ability because no differences were expected

*N* = sample size, *M* = mean, *SD* = standard deviation, *df* = degrees of freedom, *d* = Cohen's *d*

a measure of their masculine and feminine traits based on assumed gender norms (i.e., the Bem Sex Role Identity Scale; Bem, 1974). The three items concerning masculinity were then averaged to create a masculinity composite, with high scores reflecting greater self-perceived masculinity. All participants answered at least two of the three items. Cronbach's alphas were 0.89 for women and 0.78 for men. These statistics are highly similar to internal consistency estimates in previous work (Johnson et al., 1996; Storms, 1979).

In exploratory analyses, and for comparability with other recent work (Beltz, 2018; Gülgöz et al., 2019), a second bipolar masculinity score was computed. Specifically, the feminine items were reverse-coded and averaged with the masculinity items to create a single dimension with high scores reflecting greater masculinity (and low scores reflecting greater femininity). This unidimensional scale aligns with recent literature on gender expression being a single, bipolar continuum (Beltz et al., 2021; Castleberry, 2019). Cronbach's alphas were 0.90 for women and 0.84 for men, slightly higher than the unidimensional masculinity measure.

## Spatial Skills

**3D Mental Rotations** 3D mental rotations was assessed with a test consisting of 20 items to be completed within 10 min (Vandenberg & Kuse, 1978). For each item, participants were provided with a 2D target image of a 3D set of blocks and four response options (also 2D images of 3D sets of blocks). Participants were asked to select the two response options that are accurate rotations of the target in 3D space. Participants received a point for each correct response, providing a range of potential scores of 0 to 40. This test is widely used and has high test–retest reliability, with this and comparable measures consistently showing the expected gender difference (e.g., Jansen & Heil, 2009; Linn & Petersen, 1985; Voyer et al., 1995). Seventeen participants (nine women) were excluded from analyses of this measure due to a failure to follow instructions (e.g., selected more than two response options for an item).

**Geographical Knowledge** Geographical knowledge was assessed with the Modified Gallup Geography test (Snyder & Harris, 1996) consisting of 16 items with no time limit. Participants had to match a list of 16 places in the world to their location on a map. Participants received a point for each correct answer, providing a range of potential scores from 0 to 16. Past work using this measure highlights its unique utility, as it assesses both spatial skills and experience with spatially oriented stimuli (e.g., maps), and reports the expected gender difference (Berenbaum et al., 2012). Five women were excluded from analyses of this measure due to a failure to follow instructions.

**Identifying the True Horizontal** The ability to judge a horizontal line in reference to a plane was assessed with the Piagetian Water Level (Piaget & Inhelder, 1956) test that required participants to look at multiple drawings of tilted bottles each with a different water line and determine which picture most accurately depicts where the water line should be in reference to a flat surface (Berenbaum et al., 2012). It consists of 12 items. Participants received a point for each correct answer, providing a range of potential scores from 0 to 12. Construct validity for this widely-used task is high (Wittig & Allen, 1984), and meta-analytic evidence suggests that it shows the expected gender difference (Voyer et al., 1995). Three participants (two women) were excluded from analyses of this measure due to a failure to follow instructions or technical difficulties.

**Object Location Memory** Spatial location memory was assessed with a test that requires recalling the locations of previously seen objects on a 2D plane (Silverman & Eals, 1992; Silverman et al., 2007). Participants were shown an image of 27 objects for one minute, and they were told to focus on the contents of the screen. They were then shown another image containing the same 27 objects, with 14 of them in a different location. Participants had one minute to identify which objects were in a new location. Participants received a point for each correct answer, providing a range of potential scores from 0 to 14. Meta-analytic evidence suggests that women on average outperform men on this task (Voyer et al., 2007), and that this gender difference is widely generalizable (Silverman et al., 2007). Three participants (one woman) were excluded from analyses of this measure due to a failure to follow instructions.

## College Major STEM-ness

College major STEM-ness is the extent to which a discipline requires knowledge and concepts associated with STEM, and it has been used extensively in the extant literature (e.g., Hui & Lent, 2018; Muenks et al., 2020). Participants provided their college major, if they had declared one, as an open-ended survey response. Majors were coded on a five-point scale (Goldman & Hewitt, 1976). A score of 1 represents *fine arts* such as dance, music, and design. A score of 2 represents *humanities* such as English, French, and history. A score of 3 represents *social sciences* such as anthropology, economics, and psychology. A score of 4 represents *biological sciences* such as biology, kinesiology, and zoology. A score of 5 represents *physical sciences* such as chemistry, engineering, and physics. For situations in which majors could not be clearly coded into one category (e.g., biological engineering), half points were used. After achieving reliability (intraclass coefficient of 0.97) on a practice set of 121 comparable majors from a different

university, two independent raters coded each major (looking up details on the relevant university's website as needed). The average intraclass coefficient was 0.98, and the two ratings were averaged. In total, 86 majors were provided by participants. STEM-ness showed a negative skew (-0.56) suggesting majors overall had more STEM-ness, but not disproportionately.

### General Cognitive Ability

General cognitive ability was indexed by vocabulary, as the two constructs are highly related (Lezak et al., 2004) and do not show gender differences (Blakemore et al., 2009); this has been done and is recommended by past work on gender differences to help statistically differentiate between specific spatial skills and a general skill set (see Beltz et al., 2015; Sattler & Ryan, 2009). The Advanced Vocabulary test (Ekstrom et al., 1976) consists of two sets of 18 items, with each set to be completed in 4 min, and requires participants to select the correct synonym of a target word from five stimulus options. Participants received a point for each correct answer and lost a quarter point for each incorrect answer, providing a range of potential scores from -9 to 36.

### Data Analysis

Analyses were conducted, using SPSS 28.0, in four parts. The first two parts describe the nature of the data, the third and fourth parts tests the main study hypotheses. First, independent samples *t*-tests were used to examine the presence and size of gender differences. Based on past research (as reviewed in Beltz et al., 2020; Blakemore et al., 2009; Halpern, 2013), men were expected to score higher than women in masculinity, geographical knowledge, identifying the true horizontal, 3D mental rotations, and major STEM-ness, and women were expected to score higher than men on object location memory. Therefore, one-tailed tests were used due to the directional hypotheses stemming from a large, extant literature on gender differences in spatial skills, consistent with previous, comparable studies (e.g., Berenbaum et al., 2012, 2018; Geiser et al., 2008; Heil et al., 2018). No gender differences in general cognitive ability were expected, and therefore, a two-tailed test was used. Type I error was 0.05.

Second, correlations were used to examine the zero-order relations among all study variables, especially the four spatial tasks. Consistent with both quantitative and qualitative approaches to gender differences, relations were calculated separately for each gender.

Third, direct and indirect effects analyses in a mediation framework were conducted using the PROCESS macro (Hayes, 2013) to examine the relations among masculinity, spatial skills, and major STEM-ness with age and general cognitive ability as covariates. To address the first two study

aims, four models were run (these models have the structure of PROCESS model 8; Hayes, 2013). To address the first aim concerning the consistency of the relation between masculinity and cognition across four spatial tasks, separate models were run with masculinity as the predictor and each of the four spatial skills as an outcome. Direct effects, which control for major STEM-ness in the relations between masculinity and spatial skills, were evaluated in comparison to the zero-order correlations. To address the second aim concerning mechanisms underlying the relation between masculinity and spatial skills, the indirect effect (i.e., the extent to which the relation between masculinity and spatial tasks occurs via major STEM-ness) was estimated using bias-corrected 90% bootstrapped confidence intervals (with 5000 bootstrap samples). To address the third exploratory aim about the direction of relations between gender self-concept and gendered cognition, four models were also run with the spatial skills as the predictors and masculinity as the outcome (these models have the structure of PROCESS model 15; Hayes, 2013). Two statistics from the sets of four models in each direction were examined to provide insight into the differences in effects: variance explained in the outcome ( $R^2$ ) and the size of the direct effects for each gender (expressed as  $r$ ).

Fourth, potential gender differences were explored in the direct effects regarding the relations between masculinity and four spatial skills (while controlling for major STEM-ness) and the indirect effects concerning whether major STEM-ness explains the relations between masculinity and spatial tasks. Four further models were run with gender included as the quantitative moderator of relations between masculinity and major STEM-ness as well as masculinity and spatial skills; simple slope (i.e., direct effects) and indirect effects were examined separately for men and women. Based on the extant literature, there is substantial reason to question whether the sex-role hypothesis applies to women and men equally. Because the extant literature on this topic is limited and inconclusive, there is notable uncertainty whether a potential gender difference is quantitative (i.e., average gender differences on the same metric, according to Becker & Koob, 2016) or qualitative (i.e., gender differences in patterns, according to Becker & Koob, 2016). To encourage future research on this topic and to address the fourth exploratory aim about gender differences, we have included the results of these follow-up analyses that include gender as a potential moderator; please see the subsection *Gender as a Moderator to the Indirect Effects* in the online supplement.

## Results

### Gender Differences

Gender differences in all study variables are reported in Table 1, along with the means and standard deviations

separately for men and women. As expected, and compared to women, men reported greater masculinity, had higher scores on three spatial measures (i.e., 3D mental rotations, geographical knowledge, and identifying the true horizontal), and enrolled in college majors with a higher degree of STEM-ness. Unexpectedly, women did not significantly outperform men on object location memory, although the difference was in the anticipated direction. As expected, there was no significant gender difference in general cognitive ability.

### Correlations Among Study Variables

Table 2 shows the correlations among study variables, separately for each gender, with values for men below the diagonal, and values for women above the diagonal. All cognitive variables were substantially positively inter-related for women, except for some relations with object location memory and geographical knowledge which were less substantial. This was not the case for men, who had sporadic substantial links, although all were positive. Importantly, most links were small-to-moderate, suggesting some meaningful overlap, but also distinctness among the constructs, including the different measures of spatial skill. Correlations between masculinity, spatial skills, and major STEM-ness are discussed in the context of the direct and indirect effects analyses below.

### Direct and Indirect Effects Analyses

Figures 1 and 2 present results of the indirect effect analyses that contain estimates of the direct (controlling for major STEM-ness) and indirect effects (occurring via major STEM-ness). Results for each spatial skill are presented in the following format below. First, models with spatial skills as the outcome are presented (Fig. 1). Results of the major STEM-ness regression model (i.e., masculinity predicting major STEM-ness) and then the outcome regression model (i.e., masculinity and major STEM-ness predicting

the spatial skill) are described; both regressions contain the covariates of age and general cognitive ability. For the outcome models, between 1 and 14% of the variation in spatial skills was explained. Second, exploratory models with masculinity as the outcome are then presented in a parallel fashion (Fig. 2). In those outcome models, between 10 and 17% of the variance in masculinity was explained. Direct effects (absolute values, or  $r$ 's), for both sets of models, included masculinity and a spatial skill; they ranged from 0.24 for 3D mental rotations, 0.18 for geographical knowledge and identifying the true horizontal, and 0.004 for object location memory.

Sensitivity power analyses conducted in G\*Power (Faul et al., 2007) suggest that the sample size was well-powered to detect the  $R^2$  of the outcome models for seven of the eight analyses (2 predictors, 2 covariates;  $\alpha=0.05$ ;  $1-\beta=0.80$ ).

### 3D Mental Rotations

The major STEM-ness model (which included covariates and the main effect of masculinity) was significant,  $F(3, 318) = 5.60$ ,  $p < .001$ , and masculinity was a significant predictor of major STEM-ness; coefficients of key relations are seen in Fig. 1. Age was a significant covariate, such that older participants reported less major STEM-ness. The outcome model predicting 3D mental rotations (with covariates and the main effect of masculinity and major STEM-ness) was also significant,  $F(4, 317) = 12.50$ ,  $p < .001$ . Major STEM-ness was a significant predictor, and controlling for major STEM-ness, so was masculinity (i.e., the direct effect); see Fig. 1. General cognitive ability was a significant positive covariate of 3D mental rotations, as expected. Importantly, there was also an indirect effect of STEM-ness on the masculinity–3D mental rotations relation (i.e., CI does not include 0).

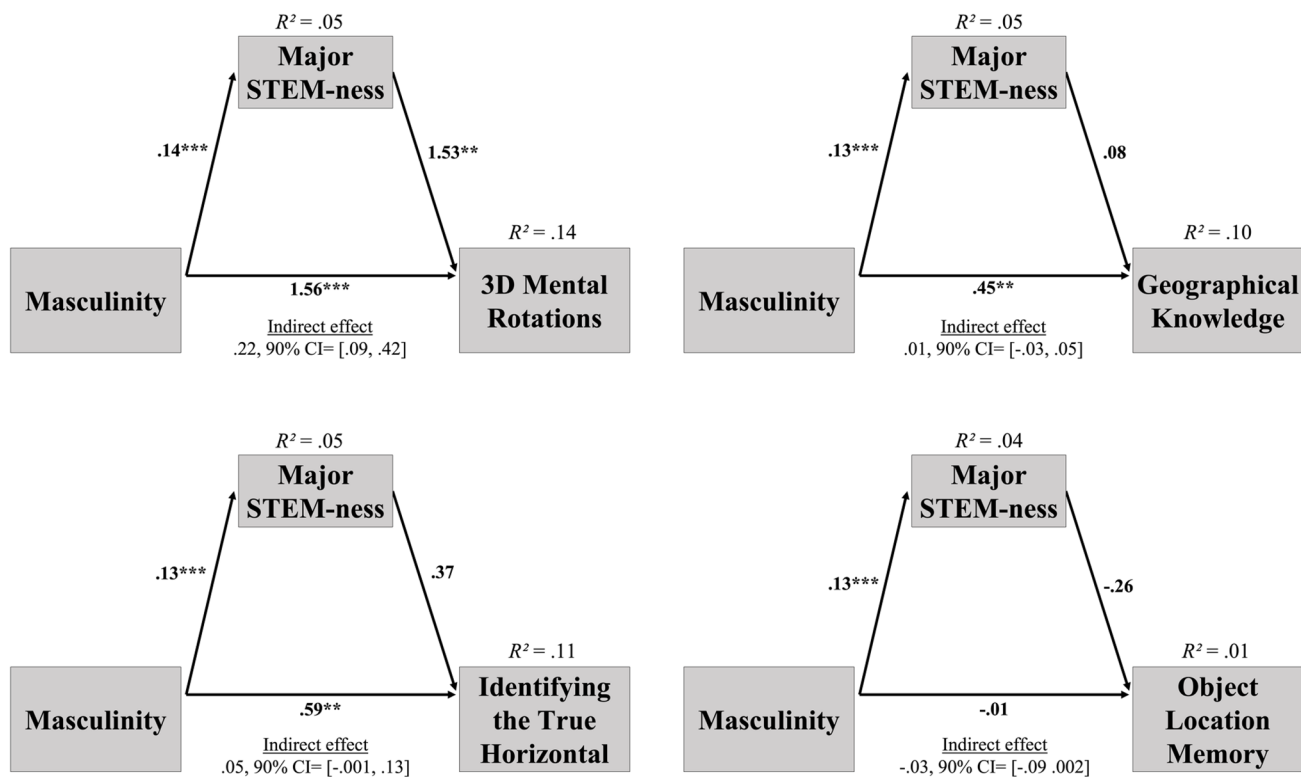
In the exploratory direction, the major STEM-ness model (which included covariates and the main effect of 3D mental rotations) was also significant,  $F(3, 318) = 6.67$ ,  $p < .001$ ,

**Table 2** Correlations Among Study Variables by Gender

	Masculinity	3D mental rotations	Geographical knowledge	Identifying the true horizontal	Object location memory	Major STEM-ness	General cognitive ability
Masculinity	–	.15*	.16*	.15*	.10	.13	.07
3D mental rotations	–.02	–	.18	.30***	.24***	.09	.22**
Geographical knowledge	–.19*	.18	–	.10	.06	–.02	.23**
Identifying the true horizontal	–.18	.30**	.17	–	.07	.07	.28***
Object location memory	–.02	.08	.06	–.17	–	–.14*	.07
Major STEM-ness	.03	.30**	.12	.13	.06	–	–.01
General cognitive ability	–.22*	.05	.28**	.24*	.07	.07	–

Correlations for men are below the diagonal, correlations for women are above the diagonal

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$



**Fig. 1** Representations of the Four Direct and Indirect Effect Analyses Conducted with Spatial Skills as Outcomes, Masculinity as the Predictor, Major STEM-ness as the Indirect Effect and General Cognitive Ability and Age as Covariates. Direct effects express the relations

between masculinity and spatial skills controlling for major STEM-ness. Indirect effects express the extent to which the relation between masculinity and spatial skills occurs via major STEM-ness. Unstandardised coefficients are shown; \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

and 3D mental rotations was a significant predictor of major STEM-ness; coefficients of this and other key relations are seen in Fig. 2. Age was a significant covariate. The model predicting masculinity (with covariates and the main effects of 3D mental rotations and major STEM-ness) was also significant,  $F(4, 317) = 15.99$ ,  $p < .001$ . Major STEM-ness was a significant predictor, and controlling for major STEM-ness, 3D mental rotations was also a significant predictor; see Fig. 2. Age was a significant covariate for masculinity such that older participants reported greater masculinity. There was also an indirect effect of STEM-ness on the 3D mental rotations-masculinity relation.

### Geographical Knowledge

The major STEM-ness model was significant,  $F(3, 330) = 5.42$ ,  $p < .01$  and masculinity was a significant predictor (Fig. 1). Age was a significant covariate. The model predicting geographical knowledge was also significant,  $F(4, 329) = 8.94$ ,  $p < .001$ . Major STEM-ness was not a significant predictor, but controlling for STEM-ness, masculinity was a significant predictor. General cognitive ability was a significant covariate of geographical knowledge. There was no

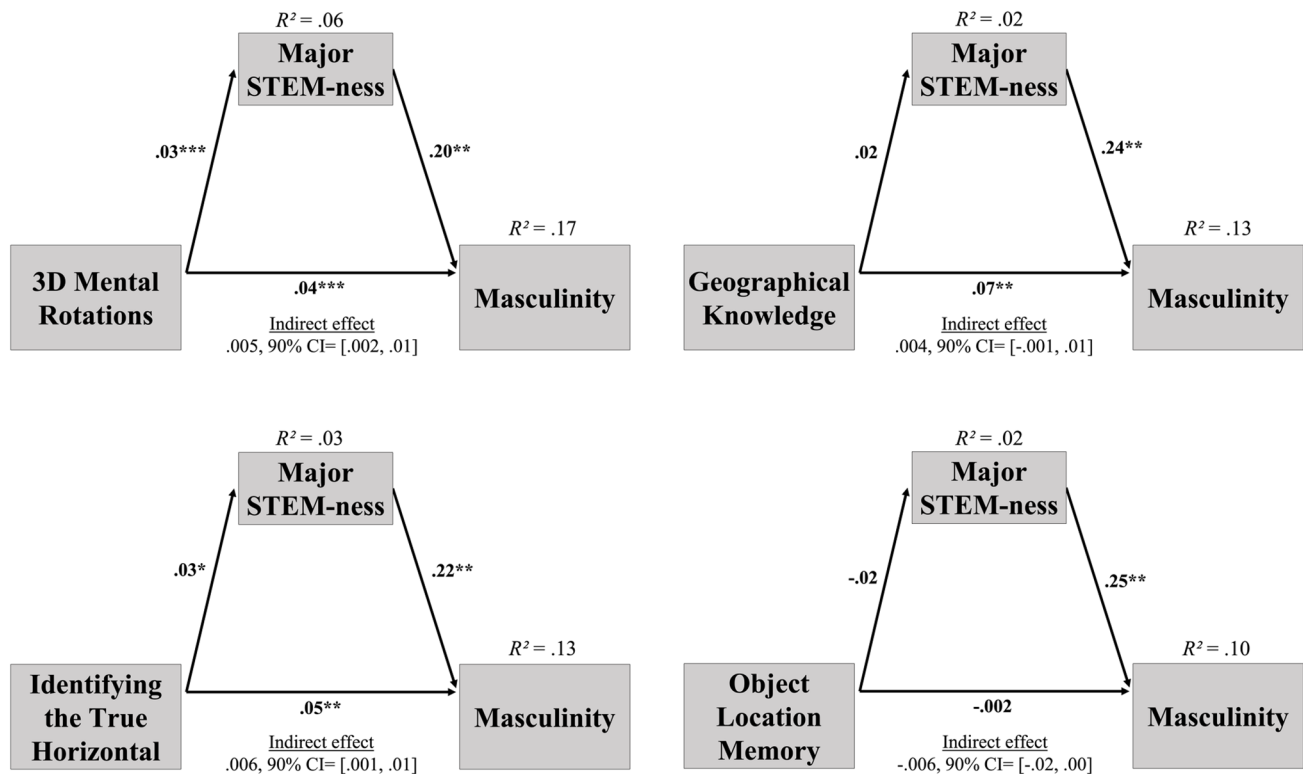
indirect effect of STEM-ness on the masculinity-geographical knowledge relation (i.e., CI does include 0).

In the exploratory direction, the major STEM-ness model was not significant,  $F(3, 330) = 1.92$ ,  $p > .05$ , and geographical knowledge was not a significant predictor of major STEM-ness either (Fig. 2). Age was a significant covariate. The model predicting masculinity was significant, though,  $F(4, 329) = 12.24$ ,  $p < .001$ . Major STEM-ness was not a significant predictor, but controlling for major STEM-ness, geographical knowledge was a significant predictor. Age was a significant covariate for masculinity. There was no indirect effect of STEM-ness on the geographical knowledge-masculinity relation.

### Identifying the True Horizontal

The major STEM-ness model was significant,  $F(3, 332) = 5.25$ ,  $p < .01$ , and masculinity was a significant predictor (Fig. 1). Age was a significant covariate. The model predicting identifying the true horizontal was also significant,  $F(4, 331) = 10.69$ ,  $p < .001$ . Major STEM-ness was not a significant predictor, but controlling for STEM-ness, masculinity was a significant predictor. General cognitive





**Fig. 2** Representations of the Four Direct and Indirect Effects Analyses Conducted with Masculinity as the Outcome, Spatial Skills as Predictors, Major STEM-ness as the Indirect Effect and General Cognitive Ability and Age as the Covariates. Direct effects express the relations

between spatial skills and masculinity controlling for major STEM-ness. Indirect effects express the extent to which the relation between spatial skills and masculinity occurs via major STEM-ness. Unstandardised coefficients are shown; \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

ability was a significant covariate. There was no indirect effect of STEM-ness on the masculinity–identifying the true horizontal relation.

In the exploratory direction, the major STEM-ness model was significant,  $F(3, 332) = 2.93$ ,  $p < .05$ , and identifying the true horizontal was a significant predictor (Fig. 2). Age was a significant covariate. The model predicting masculinity was also significant,  $F(4, 331) = 12.52$ ,  $p < .001$ . Major STEM-ness was a significant predictor, and controlling for major STEM-ness, identifying the true horizontal was also a significant predictor. Age was a significant covariate for masculinity. There was also an indirect effect of STEM-ness on the identifying the true horizontal–masculinity relation.

### Object Location Memory

The major STEM-ness model was significant,  $F(3, 332) = 5.14$ ,  $p < .01$  and masculinity was a significant predictor (Fig. 1). Age was a significant covariate. The model predicting object location memory was not significant,  $F(4, 331) = 1.10$ ,  $p > .05$ . Major STEM-ness was not a significant predictor, and controlling for STEM-ness, neither was masculinity. There were no significant covariates for object

location memory. There was no indirect effect of STEM-ness on the masculinity–object location memory relation.

In the exploratory direction, the major STEM-ness model was not significant,  $F(3, 332) = 2.04$ ,  $p > .05$ , and object location memory was not a significant predictor (Fig. 2). Age was a significant covariate. The model predicting masculinity was significant,  $F(4, 331) = 9.41$ ,  $p < .001$ . Major STEM-ness was a significant predictor, but controlling for major STEM-ness, object location memory was not. Age was a significant covariate of masculinity. There was also no indirect effect of STEM-ness on the object location memory–masculinity relation.

## Summary of Supplementary Analyses

### Masculinity as a Bipolar Measure

The aforementioned analyses focused on the primary unidimensional measure of masculinity (in which low scores reflect low masculinity and high scores reflect high masculinity). Parallel exploratory analyses for the bipolar measure of masculinity (in which low scores reflect femininity and high scores reflect masculinity) were also conducted and are

reported in the *Masculinity as a Bipolar Measure* subsection of the online supplement. The pattern of results between the two sets of analyses are broadly the same, with no notable differences.

### Gender as a Moderator

To address the fourth aim of the study and to examine whether the pattern of results depended upon gender, moderated indirect effect analyses were conducted, and reported in the *Gender as a Moderator to the Indirect Effects* subsection of the online supplement. These analyses included gender as a moderator of the relations between masculinity and major STEM-ness and between major STEM-ness and spatial skills. These analyses are consistent with a qualitative gender difference: The test of moderated mediation was not significant for any model, but significant direct effects between masculinity and spatial skills were consistently present and some indirect effects were occasionally present, but all only for women.

The results of analyses which combine the two aforementioned approaches (i.e., gender is included as a moderator with masculinity as a bipolar measure) are available from the authors on request. Results were similar: there were significant, small-to-medium-sized relations between masculinity and spatial skills which were significant only for women (instead of both genders), and major STEM-ness accounted for a significant amount of variation between masculinity and 3D mental rotations and identifying the true horizontal (for models in the exploratory direction), but again only for women – tests of moderated mediation suggested a lack of significant gender differences in the indirect effects.

## Discussion

The goal of this study was to examine the consistency of the relation between gender self-concept (indexed by self-perceived masculinity) and gendered cognition (indexed by a battery of spatial tasks: 3D mental rotations, geographical knowledge, identifying the true horizontal and object location memory) and to determine whether this relation was undergirded by gendered activities, experiences, and interests (indexed by college major STEM-ness). The direction of effects (i.e., whether masculinity predicts spatial skills or spatial skills predict masculinity) and the presence of gender differences was also explored. The goal was achieved through a series of direct and indirect effects analyses and applied to a large cross-sectional data set. Results provided partial support for the sex-role mediation hypothesis. Although they revealed significant, small-to-medium-sized relations between masculinity and spatial skills, exploratory analyses suggested there is statistical and conceptual

utility in considering relations from spatial skills to masculinity. Results also expanded upon the sex-role mediation hypothesis by revealing that major STEM-ness accounted for a significant amount of variation between masculinity and 3D mental rotations and identifying the true horizontal.

Inferences about the consistency of relations between masculinity and spatial skills were based on whether bivariate correlations and direct effects (controlling for major STEM-ness) were present for the four spatial skills measured in this study, extending previous work which only included 3D mental rotations and identifying the true horizontal of these four skills (Reilly et al., 2016). Direct effects were significant for 3D mental rotations, geographical knowledge and identifying the true horizontal. Differences among spatial skills emphasise that spatial ability is not unitary, but rather, consists of several overlapping, yet distinct constructs that are multidimensional and multidetermined (Newcombe, 2002).

Inferences about whether relations between masculinity and spatial skills occur via gendered activities, experiences, and interests were based on indirect effects of college major STEM-ness. Indirect effects were detected for 3D mental rotations and identifying the true horizontal (although only in an exploratory alternative direction for the latter spatial skill). Interestingly, these are the two ‘purest’ spatial skills in this study, as they rely primarily on spatial visualization, mentalization, and orientation and are not confounded by general knowledge or short-term memory like geographical knowledge and object location memory, respectively (again demonstrating the multidetermined nature of spatial skills). Given the positive relation between interest in science and spatial skills (Lubinski, 2010), it follows that gendered activities, experiences, and interests as indexed by major STEM-ness would disproportionately be related to the ‘purest’ spatial skills. It should be noted that direct effects (which control for major STEM-ness) were also significant for all spatial skills except object location memory. This suggests that, even for the ‘purest’ spatial skills that showed indirect effects, that there is an effect between masculinity and spatial skills beyond what college major STEM-ness reflects in gendered activities, experiences, and interests.

Exploratory analyses about directionality were based on variation explained in the outcomes as well as the presence of direct effects in eight models – four linked to the sex-role hypothesis that had spatial skills as outcomes and four novel models with masculinity as the outcome. Meaningful variation was explained in models with masculinity as the outcome as well as in models with spatial skills as outcomes, suggesting that the former models are also worthy of future study. This direction of effects counters the sex-role hypothesis, but it is by no means unlikely or unsupported by the theoretical and empirical literature. In fact, Nash (1979) even suggested developmental mechanisms that may underlie

such effects: “as children begin to evaluate their competencies, it becomes apparent to them that their talents are either congruent or incongruent with their gender” (p. 291). This is consistent with a cognitive constructionist approach to gender self-concept (Martin & Halverson, 1981), acknowledging that individuals both affect and are affected by their gendered behaviors (Liben & Bigler, 2002). Indeed, it has long been posited that gender self-concept arises from a ‘complex calculus’ of individuals’ awareness of how congruent their behaviors are with gendered norms and how salient those gendered behaviours are to their sense of self (Spence, 1985).

Although long-standing theories of gender development support statistical inferences from this study suggesting that spatial skills *lead to* masculine self-concepts, study inferences were based on statistical estimation of indirect effects (and not formal mediation; see Preacher & Hayes, 2008) in cross-sectional data, so causality cannot be claimed. Longitudinal work is needed across childhood and adolescence and even into adulthood to help inform quasi-causal inferences about the direction of the relation between masculinity and spatial skills. Key questions for future work concern the sample ages and the repeated assessment schedule that would best examine changes in these constructs, as it is possible (even likely) that effects change over time. Nonetheless, these exploratory analyses are still valuable: significant effects in models of masculinity – at a minimum – indicate that the direction of the sex-role hypothesis is debated and requires further investigation.

It is notable that there were only direct and indirect effects present for models that included spatial skills that – on average – show a masculine advantage. This poses an interesting question about quantitative and qualitative gender differences – a quantitative difference would reflect different magnitudes of the same relations among masculinity, cognition and STEM-ness between women and men, whereas a qualitative difference would reflect that the processes underlying inter-relations among masculinity, cognition and STEM-ness are gendered, acknowledging that the same pattern of relations may not be applicable to women and men. Given that masculinity, major STEM-ness, and most spatial skills typically show gender differences favoring men, a further question to be explored is whether it matters if these variables are gender-congruent or gender-incongruent, depending on one’s gender (Spence, 1985), and whether qualitative gender differences underlie this finding.

The potential for gender differences in the direct and indirect effects was also further explored (presented in the *Gender as a Moderator to the Indirect Effects* subsection of the online supplement). Despite the lack of quantitative gender differences (i.e., a lack of significant moderated mediations), exploratory findings are consistent with qualitative gender differences, which suggest processes underlying masculinity,

STEM-ness, and spatial skills differ by gender. This could be explicitly examined in future research, particularly developmental work. For example, evidence suggests that women in STEM programs scored higher on mental rotations tasks than women not in STEM programs only when the former group preferred spatial toys in childhood (Moè et al., 2018), suggesting potential self-selection into STEM careers based on likely multidetermined characteristics, which could begin to be captured with longitudinal data on antecedents and outcomes of spatial skills.

### Limitations and Future Research Directions

The results and conclusions of this study must be considered in the context of the study design, analyses conducted, and limitations of the dataset. Strengths of this study include the large sample size of a narrow age range compared to the extant literature (standard deviation of 0.96 years compared to 8.03 in Reilly et al., 2016), the inclusion of a feminine spatial measure (i.e., object location memory), and the statistical discernment between general cognitive abilities and specific spatial skills. There are also some limitations.

First, and as previously noted, the current study was cross-sectional. Estimates of indirect effects in cross-sectional data can be biased (Maxwell & Cole, 2007), but this study’s conclusions rely on bias-corrected bootstrapped confidence intervals of the indirect effects, which is preferable to and is more robust than relying on statistical significance (or not) of relations (Mackinnon et al., 2004; Maxwell & Cole, 2007). Thus, exploratory inferences about directionality are statistical and suggestive, and future confirmatory longitudinal work is needed. This longitudinal work could qualify the inferences drawn here about young adults. For example, significant direct or indirect effects between spatial skills and masculinity might be present for adolescent boys whose identities are still forming. Still, studying college students provided the unique opportunity to consider the effects of college major STEM-ness (unconfounded workforce experiences) in a sample with a limited age range.

Second, although this study had the largest sample size of single studies on this topic to-date, it still may have been too small to detect some effects. For instance, women outperformed men in object location memory with an effect size of  $d=0.19$ , but the difference was not significant. Expected gender differences (or the lack thereof in the case of general cognitive ability) were found in all other study variables (Beltz et al., 2020; Halpern, 2013). Also, there was a consistent pattern of negative direct effects for men (mirroring the consistent positive direct effects for women), but none were significant; indirect effects also did not statistically differ between men and women. This could reflect that these effects are qualitative more than quantitative, but methodological explanations for the patterns of gendered effects are

also possible; because there were fewer men than women in this sample, there is reduced power to detect effects in men compared to women and to detect interactions in the indirect effects (i.e., the moderated indirect effects analyses).

Third, there was a significant age difference between men and women ( $d = 0.55$ ). Although this is a medium-sized effect (Cohen, 1988), it may be due to the limited age range of the sample. Indeed, it only reflects a six-month age difference during which significant changes in masculinity or spatial skills would not be expected. Regardless, in these models, age was consistently inversely related to major STEM-ness, which aligns with research showing STEM interests decrease over adolescence, particularly for girls (George, 2006; Sadler et al., 2012). Age was also positively related to masculinity in some models, which is consistent with small effects seen in the literature (Barrett & Raskin White, 2002). Future work should utilize samples matched on age to buttress inferences.

Fourth, there is variety in how masculinity is operationalised in studies of gender self-concept and questions about how that operationalization might affect results and inferences. Although studies on the sex-role hypothesis have measured gender self-concept by the desire to be the opposite gender (e.g., Newcombe & Dubas, 1992) and identification with gendered personality traits (typically with the Bem Sex Role Inventory; Bem, 1974), a measure of gender self-expression was utilized in this study (i.e., Storms, 1979). Different aspects of gender self-concept (i.e., gender expression and gendered personality qualities) are correlated (Hyde et al., 2019; Ruble et al., 2006; Spence & Buckner, 1995), so it is possible that all of these operationalizations reflect related latent constructs, and that this study's focus on gender expression yielded unique relations. Future work comparing links between multiple spatial skills and multiple operationalizations of masculinity would be quite informative.

Fifth, the primary analyses reported here concerned masculinity as a unidimensional construct, but it is important to note that conceptualizations of gender expression (especially as it is related to manifestations of gender identity as a continuum) warrant consideration of masculinity as one pole of a bipolar dimension, and femininity as the other pole. Indeed, refining the argument (Spence, 1984, 1985) that gender expression and gendered traits are related but not interchangeable elements of a multifactorial understanding of gender, Eagly and Wood (2017) argue that there is a strong relation between how an individual perceives gendered traits and how they assess their own masculinity and femininity. For that reason, parallel analyses using a bipolar measure of masculinity were conducted (presented in the *Masculinity as a Bipolar Measure* subsection of the online supplement). As described earlier, there were no notable differences in the patterns of results between the two sets of analyses,

potentially suggesting that gender expression is a continuum at least in some circumstances or samples. If femininity were a completely independent dimension of gender expression in the context of this study, then its inclusion in the bipolar subscale with masculinity would have drastically altered the pattern of results, but it did not. Therefore, both conceptualizations of masculinity likely hold at least some relevance for examinations of gender expression.

Sixth, although tests of the moderated indirect effects (presented in the *Gender as a Moderator to the Indirect Effects* subsection of the online supplement) suggested that the effects were not significantly different for women and men (even though they were greater than 0 in women, but not in men), analyses that included gender as a moderator may reflect that gender differences in associations among masculinity, major STEM-ness, and spatial skills are not (or are not just) quantitative, but rather, qualitative (see Becker & Koob, 2016; Beltz et al., 2019). Although some have suggested there are stronger links between masculinity and spatial skills in girls/women than in boys/men (Signorella & Jamison, 1986), others have reported overall larger links for boys/men than girls/women (Reilly & Neumann, 2013). If gender differences in these processes are truly quantitative, then this might reflect a lack of statistical power even in this large sample or restricted range of masculinity scores, as there was little overlap in scores between men and women. Thus, future work in larger, more diverse samples is needed.

## Practice Implications

The premise of this study and its central findings concerning links between masculinity and spatial skills via college major STEM-ness is relevant to potent conversations surrounding gender disparities in STEM. Specifically, findings are consistent with research highlighting that women choose non-STEM career paths for reasons other than ability, such as interests, lifestyle and career values, and the perceived communal orientation of the work (e.g., Boucher et al., 2017; Eccles & Wang, 2016; Wang et al., 2013; Williams & Ceci, 2015): self-perceived gender self-concept as assessed in this study may overlap with some of these gendered influences, as well as have unique links with STEM orientation and spatial skills. That is an exciting avenue for future, applied research. Such research should be longitudinal to consider directionality of influences and it should gauge participants' perceptions of the gender normativity and saliency of their college major or careers (which can potentially be affected by practice or training). In combination with such subsequent experimental and longitudinal work, there is the potential for the findings of this study to help inform practice-oriented research aimed at promoting gender equality in STEM.

## Conclusion

“It is crucial that a theory which proposes to link any aspect of sex-role concept with gender-related differences in cognitive performance be able to explain why boys do better on spatial-quantitative tests” (Nash, 1979, p. 264). The sex-role hypothesis suggests that, in both men and women, heightened masculinity leads to enhanced spatial skills via gendered activities, experiences, and interests. Although there is some empirical support for the direct link between masculinity and spatial skills in the extant literature, the current study for the first time examined STEM education as a gendered mechanism underlying the link. As expected, significant correlations between masculinity and spatial skills were present for all spatial skills which typically show a masculine advantage (3D mental rotations, geographical knowledge and identifying the true horizontal). Importantly, indirect effects analyses suggested that, for select spatial skills, the STEM-ness of students’ college majors partially accounted for those direct relations. Moreover, exploratory analyses tentatively suggested that the effects may be more pronounced in women than men, and they call into question the direction of the sex-role hypothesis, as statistical links from spatial skills to masculinity were also significant, in line with a dual pathways approach to gender differentiation. Together, these findings highlight that gender is a multifaceted system of constructs, with complex patterns of individual differences. Although these findings require replication and extension in future longitudinal studies, they nonetheless highlight that there is a relatively consistent and compelling link between self-perceived masculinity and spatial skills that, at least for 3D mental rotations and identifying the true horizontal, could be due to exposure to gendered activities, experiences, and interests, such as through STEM education. Potentially, this line of research could one day inform policy and interventions to reduce gender disparities and increase equity in participation and success in STEM.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11199-022-01308-y>.

**Acknowledgements** The authors thank Dr. Sheri Berenbaum for her help designing the original study. They also thank research assistants in the Berenbaum Lab at the Pennsylvania State University and the Methods, Sex differences and Development Lab at the University of Michigan for their help with data collection, management, and coding. Finally, they thank all of the participants who made this research possible.

**Authors’ Contributions** Study design, material preparation, and data collection were performed by Adriene Beltz. The idea for this manuscript was conceived by Dominic Kelly. Analyses were performed by Dominic Kelly and Adriene Beltz. The manuscript was written by Dominic Kelly with critical input from Adriene Beltz. Both authors read and approved the submitted manuscript.

**Funding** Dominic Kelly was supported by the US-UK Fulbright Commission; Adriene Beltz was supported by the Jacobs Foundation.

**Availability of Data and Material** None.

**Code Availability** None.

## Declarations

**Ethics Approval** This study was approved by the Institutional Review Boards of the Pennsylvania State University and the University of Michigan. All of the procedures in this study are performed in line with the principles of the Declaration of Helsinki.

**Consent to Participate and Publish** All participants provided informed implied consent to participate in this study and for the results to be published.

**Conflicts of Interest / Competing Interests** None.

## References

- Andersen, L. (2014). Visual-spatial ability: Important in STEM, ignored in gifted education. *Roeper Review*, 36(2), 114–121. <https://doi.org/10.1080/02783193.2014.884198>
- Barrett, A. E., & Raskin White, H. (2002). Trajectories of gender role orientations in adolescence and early adulthood: A prospective study of the mental health effects of masculinity and femininity. *Journal of Health and Social Behavior*, 43(4), 451–468. <https://doi.org/10.2307/3090237>
- Becker, J. B., & Koob, G. F. (2016). Sex differences in animal models: Focus on addiction. *Pharmacological Reviews*, 68(2), 242–263. <https://doi.org/10.1124/pr.115.011163>
- Beltz, A. M. (2018). Gendered mechanisms underlie the relation between pubertal timing and adult depressive symptoms. *Journal of Adolescent Health*, 62(6), 722–728. <https://doi.org/10.1016/j.jadohealth.2017.12.019>
- Beltz, A. M., Beery, A. K., & Becker, J. B. (2019). Analysis of sex differences in pre-clinical and clinical data sets. *Neuropsychopharmacology*, 44(13), 2155–2158. <https://doi.org/10.1038/s41386-019-0524-3>
- Beltz, A. M., Hampson, E., & Berenbaum, S. A. (2015). Oral contraceptives and cognition: A role for ethinyl estradiol. *Hormones and Behavior*, 74, 209–217. <https://doi.org/10.1016/j.yhbeh.2015.06.012>
- Beltz, A. M., Kelly, D. P., & Berenbaum, S. A. (2020). Sex differences in brain and behavioral development. In J. Rubenstein & P. Rakic (Eds.), *Neural circuit and cognitive development* (2nd ed., pp. 585–638). <https://doi.org/10.1016/b978-0-12-814411-4.00027-5>
- Beltz, A. M., Loviska, A. M., & Weigard, A. (2021). Daily gender expression is associated with psychological adjustment for some people, but mainly men. *Scientific Reports*, 11(1), 1–10. <https://doi.org/10.1038/s41598-021-88279-4>
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42(2), 155–162. <https://www.ncbi.nlm.nih.gov/pubmed/4823550>
- Benenson, J. F., Tennyson, R., & Wrangham, R. W. (2011). Male more than female infants imitate propulsive motion. *Cognition*, 121(2), 262–267. <https://doi.org/10.1016/j.cognition.2011.07.006>
- Berenbaum, S. A., Beltz, A. M., Bryk, K., & McHale, S. (2018). Gendered peer involvement in girls with Congenital Adrenal Hyperplasia: Effects of prenatal androgens, gendered activities,

- and gender cognitions. *Archives of Sexual Behavior*, 47(4), 915–929. <https://doi.org/10.1007/s10508-017-1112-4>
- Berenbaum, S. A., Bryk, K. L., & Beltz, A. M. (2012). Early androgen effects on spatial and mechanical abilities: Evidence from congenital adrenal hyperplasia. *Behavioral Neuroscience*, 126(1), 86–96. <https://doi.org/10.1037/a0026652>
- Blakemore, J. E. O., Berenbaum, S. A., & Liben, L. S. (2009). Motor development and cognition. In *Gender development* (pp. 69–94). Psychology Press.
- Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1005.
- Boucher, K. L., Fuesting, M. A., Diekman, A. B., & Murphy, M. C. (2017). Can I work with and help others in this field? How communal goals influence interest and participation in STEM fields. *Frontiers in Psychology*, 8, 901. <https://doi.org/10.3389/fpsyg.2017.00901>
- Butcher, J., Dahlstrom, W., Graham, J., Tellegen, A., & Kaemmer, B. (1989). *Manual for administering and scoring the MMPI-2*. University of Minnesota Press.
- Castleberry, J. (2019). Addressing the gender continuum: A concept analysis. *Journal of Transcultural Nursing*, 30(4), 403–409. <https://doi.org/10.1177/1043659618818722>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Di Dio, L., Saragovi, C., Koestner, R., & Aubé, J. (1996). Linking personal values to gender. *Sex Roles*, 34(9), 621–636. <https://doi.org/10.1007/bf01551498>
- DiDonato, M. D., & Berenbaum, S. A. (2011). The benefits and drawbacks of gender typing: How different dimensions are related to psychological adjustment. *Archives of Sexual Behavior*, 40(2), 457–463. <https://doi.org/10.1007/s10508-010-9620-5>
- Eagly, A. H., & Wood, W. (2017). Janet Taylor Spence: Innovator in the study of gender. *Sex Roles*, 77(11), 725–733. <https://doi.org/10.1007/s11199-017-0835-y>
- Eccles, J. S., & Wang, M.-T. (2016). What motivates females and males to pursue careers in mathematics and science? *International Journal of Behavioral Development*, 40(2), 100–106. <https://doi.org/10.1177/0165025415616201>
- Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Kit of factor-referenced cognitive tests*. Educational Testing Service.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence*, 36(6), 556–563. <https://doi.org/10.1016/j.intell.2007.12.003>
- George, R. (2006). A cross-domain analysis of change in students' attitudes toward science and attitudes about the utility of science. *International Journal of Science Education*, 28(6), 571–589. <https://doi.org/10.1080/09500690500338755>
- Goldman, R. D., & Hewitt, B. N. (1976). The Scholastic Aptitude Test “explains” why college men major in science more often than college women. *Journal of Counseling Psychology*, 23(1), 50. <https://doi.org/10.1037/0022-0167.23.1.50>
- Gülgöz, S., Glazier, J. J., Enright, E. A., Alonso, D. J., Durwood, L. J., Fast, A. A., Lowe, R., Ji, C., Heer, J., Martin, C. L., & Olson, K. R. (2019). Similarity in transgender and cisgender children's gender development. *Proceedings of the National Academy of Sciences*, 116(49), 24480–24485. <https://doi.org/10.1073/pnas.1909367116>
- Halpern, D. F. (2013). *Sex differences in cognitive abilities* (4th ed.). Psychology Press. <https://doi.org/10.4324/9780203816530>
- Halpern, D. F., Straight, C. A., & Stephenson, C. L. (2011). Beliefs about cognitive gender differences: Accurate for direction, underestimated for size. *Sex Roles*, 64(5), 336–347. <https://doi.org/10.1007/s11199-010-9891-2>
- Hayes, A. F. (2013). *Introduction to mediation, moderation and conditional process analysis: A regression based approach*. Guilford Press.
- Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin and Review*, 25(3), 1212–1219. <https://doi.org/10.3758/s13423-017-1347-z>
- Heil, M., Krüger, M., Krist, H., Johnson, S. P., & Moore, D. S. (2018). Adults' sex difference in a dynamic mental rotation task. *Journal of Individual Differences*, 39(1), 48–52. <https://doi.org/10.1027/1614-0001/a000248>
- Hui, K., & Lent, R. W. (2018). The roles of family, culture, and social cognitive variables in the career interests and goals of Asian American college students. *Journal of Counseling Psychology*, 65(1), 98–109. <https://doi.org/10.1037/cou0000235>
- Hungerford, J. K., & Sobolew-Shubin, A. P. (1987). Sex-role identity, gender identity, and self-schemata. *Psychology of Women Quarterly*, 11(1), 1–10. <https://doi.org/10.1111/j.1471-6402.1987.tb00769.x>
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581–592. <https://doi.org/10.1037/0003-066X.60.6.581>
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74(2), 171–193. <https://doi.org/10.1037/amp0000307>
- Jansen, P., & Heil, M. (2009). Gender differences in mental rotation across adulthood. *Experimental Aging Research*, 36(1), 94–104. <https://doi.org/10.1080/03610730903422762>
- Johnson, M. E., Jones, G., & Brems, C. (1996). Concurrent validity of the MMPI-2 feminine gender role (GF) and masculine gender role (GM) scales. *Journal of Personality Assessment*, 66(1), 153–168. [https://doi.org/10.1207/s15327752jpa6601\\_12](https://doi.org/10.1207/s15327752jpa6601_12)
- Johnson, W., & Bouchard, T., Jr. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33(4), 393–416. <https://doi.org/10.1016/j.intell.2004.12.002>
- Lehavot, K., King, K. M., & Simoni, J. M. (2011). Development and validation of a gender expression measure among sexual minority women. *Psychology of Women Quarterly*, 35(3), 381–400. <https://doi.org/10.1177/0361684311413554>
- Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. Oxford University Press.
- Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, 67(2), 1–147. <https://doi.org/10.1111/1540-5834.t01-1-00187>
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56(6), 1479–1498. <https://doi.org/10.2307/1130467>
- Lippa, R. A. (2002). Gender-related traits of heterosexual and homosexual men and women. *Archives of Sexual Behavior*, 31(1), 83–98. <https://doi.org/10.1023/a:1014035302843>
- Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, 49(4), 344–351. <https://doi.org/10.1016/j.paid.2010.03.022>
- Mackinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99. [https://doi.org/10.1207/s15327906mbr3901\\_4](https://doi.org/10.1207/s15327906mbr3901_4)

- Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, 52(4), 1119–1134. <https://doi.org/10.2307/1129498>
- Martin, H., & Finn, S. E. (2010). *Masculinity and femininity in the MMPI-2 and MMPI-A*. University of Minnesota Press.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12(1), 23–44. <https://doi.org/10.1037/1082-989X.12.1.23>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487. <https://doi.org/10.1037/a0039400>
- McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27(5), 486–493. <https://doi.org/10.1016/j.appdev.2006.06.003>
- 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>
- Muenks, K., Grossnickle Peterson, E., Green, A. E., Kolvoord, R. A., & Uttal, D. H. (2020). Parents’ beliefs about high school students’ spatial abilities: Gender differences and associations with parent encouragement to pursue a STEM career and students’ STEM career intentions. *Sex Roles*, 82, 570–583. <https://doi.org/10.1007/s11199-019-01072-6>
- Nash, S. C. (1979). Sex role as a mediator of intellectual functioning. In M. A. Wittig & A. C. Peterson (Eds.), *Sex-related differences in cognitive functioning* (pp. 263–302). Academic Press.
- Newcombe, N., Bandura, M. M., & Taylor, D. G. (1983). Sex-differences in spatial ability and spatial activities. *Sex Roles*, 9(3), 377–386. <https://doi.org/10.1007/BF00289672>
- Newcombe, N., & Dubas, J. S. (1992). A longitudinal study of predictors of spatial ability in adolescent females. *Child Development*, 63(1), 37–46. <https://doi.org/10.2307/1130899>
- Newcombe, N. S. (2002). Spatial cognition. In H. Pashler & D. Medin (Eds.), *Steven’s handbook of experimental psychology: Memory and cognitive processes* (pp. 113–163). John Wiley & Sons Inc.
- O’Heron, C. A., & Orlofsky, J. L. (1990). Stereotypic and nonstereotypic sex role trait and behavior orientations, gender identity, and psychological adjustment. *Journal of Personality and Social Psychology*, 58(1), 134–143. <https://doi.org/10.1037/0022-3514.58.1.134>
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531–536. <https://doi.org/10.1177/1745691612463401>
- Piaget, J., & Inhelder, B. (1956). *The child’s conception of space*. Routledge & Kegan Paul.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavioral Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/brm.40.3.879>
- Reilly, D., & Neumann, D. L. (2013). Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles*, 68(9–10), 521–535. <https://doi.org/10.1007/s11199-013-0269-0>
- Reilly, D., Neumann, D. L., & Andrews, G. (2016). Sex and sex-role differences in specific cognitive abilities. *Intelligence*, 54, 147–158. <https://doi.org/10.1016/j.intell.2015.12.004>
- Riley, J. G., & Jones, R. B. (2007). Review of research: When girls and boys play: What research tells us. *Childhood Education*, 84(1), 38–43. <https://doi.org/10.1080/00094056.2007.10522968>
- Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (Vol. 3, pp. 858–932). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470147658.chpsy0314>
- Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education*, 96(3), 411–427. <https://doi.org/10.1002/sce.21007>
- Sattler, J. M., & Ryan, J. J. (2009). *Assessment with the WAIS-IV*. Jerome M. Sattler.
- Signorella, M. L., & Jamison, W. (1986). Masculinity, femininity, androgyny, and cognitive performance: A meta-analysis. *Psychological Bulletin*, 100(2), 207. <https://doi.org/10.1037/0033-2909.100.2.207>
- Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of Sexual Behavior*, 36(2), 261–268. <https://doi.org/10.1007/s10508-006-9168-6>
- Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 533–549). Oxford University Press.
- Snyder, P. J., & Harris, L. J. (1996). Where in the world am I? Sex and handedness differences in knowledge of geography. *Perceptual and Motor Skills*, 82(3\_suppl), 1379–1385. <https://doi.org/10.2466/pms.1996.82.3c.1379>
- Spence, J. T. (1984). Gender identity and its implications for concepts of masculinity and femininity. In T. B. Sonderegger (Ed.), *Nebraska symposium on motivation* (Vol. 32, pp. 59–96). University of Nebraska Press.
- Spence, J. T. (1985). Implications for the concepts of masculinity and femininity. In T. B. Sonderegger (Ed.), *Nebraska symposium on motivation* (Vol. 32, pp. 59–96). University of Nebraska Press.
- Spence, J. T., & Buckner, C. (1995). Masculinity and femininity: Defining the undefinable. In P. J. Kalbfleisch & M. J. Cody (Eds.), *Gender, power, and communication in human relationships* (pp. 105–138). Lawrence Erlbaum.
- Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. University of Texas Press.
- Spence, J. T., Helmreich, R. L., & Holahan, C. K. (1979). Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. *Journal of Personality and Social Psychology*, 37(10), 1673–1682. <https://doi.org/10.1037/0022-3514.37.10.1673>
- Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). *The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity-femininity*. University of Texas.
- Steele, S. M., Belvy, L., Veldhuis, C. B., Martin, K., Nisi, R., & Hughes, T. L. (2019). Femininity, masculinity, and body image in a community-based sample of lesbian and bisexual women. *Women and Health*, 59(8), 829–844. <https://doi.org/10.1080/03630242.2019.1567645>
- Storms, M. D. (1979). Sex role identity and its relationships to sex role attributes and sex role stereotypes. *Journal of Personality and Social Psychology*, 37(10), 1779–1789. <https://doi.org/10.1037/0022-3514.37.10.1779>
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2), 599–604. <https://doi.org/10.2466/pms.1978.47.2.599>
- Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin & Review*, 18(2), 267–277. <https://doi.org/10.3758/s13423-010-0042-0>
- Voyer, D., Postma, A., Brake, B., & Imperato-McGinley, J. (2007). Gender differences in object location memory: A meta-analysis. *Psychonomic Bulletin & Review*, 14(1), 23–38. <https://doi.org/10.3758/bf03194024>

- 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*(2), 250–270. <https://doi.org/10.1037//0033-2909.117.2.250>
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835. <https://doi.org/10.1037/a0016127>
- 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*(5), 770–775. <https://doi.org/10.1177/0956797612458937>
- 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*, *112*(17), 5360–5365. <https://doi.org/10.1073/pnas.1418878112>
- Wittig, M. A., & Allen, M. J. (1984). Measurement of adult performance on Piaget's water horizontality task. *Intelligence*, *8*(4), 305–313. [https://doi.org/10.1016/0160-2896\(84\)90014-x](https://doi.org/10.1016/0160-2896(84)90014-x)
- Wood, W., & Eagly, A. H. (2015). Two traditions of research on gender identity. *Sex Roles*, *73*(11), 461–473. <https://doi.org/10.1007/s11199-015-0480-2>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.