Gender-Related Differences in Spatial Ability: A Meta-Analytic Review
Abstract

Although gender-related differences in highly gender typed cognitive abilities are of considerable interest to educators and cognitive researchers alike, relatively little progress has been made in understanding the psychological processes that lead to them. Nash (1979) proposed a sex-role mediation hypothesis for such differences, with particular emphasis on spatial ability. However, changes in gender equality and gender stereotypes in the decades since merit a re-examination of whether a gender-role association still holds (Feingold, 1988).

A meta-analysis of 12 studies that examined gender-role identity and mental rotation performance was conducted. These included studies from the United Kingdom, Canada, Poland, Croatia, and the United States of America. The mean effect size for masculinity was $r = .30$ for men and $r = .23$ for women; no association was found between femininity and mental rotation. This effect size was slightly larger than that found previously by Signorella and Jamison (1986), and exceeds many other factors known to influence spatial ability. The implications of sex-role mediation of gender differences are discussed and future research directions are identified.

Keywords: gender differences, spatial ability, sex-role mediation, gender roles, mental rotation, meta-analysis
Gender-Related Differences in Spatial Ability: A Meta-Analytic Review

Introduction

Though progress has been made in closing gaps in recent decades, women still remain underrepresented in science, technology, engineering and mathematics (STEM)-related fields in the United States with fewer women entering these fields in tertiary education (National Science Foundation, 2011). Concerns about the underrepresentation of women are also present in many other countries, including Britain (Brosnan, 1998) and Australia (Bell, 2010). Although exceptions exist for psychology and medical sciences (Hyde, 2007b), in general women are underrepresented in the sciences at a graduate level, as well scoring lower in tests of mathematics and science achievement at school within the U.S. (Gallagher & Kaufman, 2005; Hedges & Nowell, 1995). These findings are also supported by more recent reviews of mathematics and science literacy in large international assessments of student achievement such as the Programme for International Student Achievement (Else-Quest, Hyde, & Linn, 2010; Guiso, Monte, Sapienza, & Zingales, 2008; Reilly, 2012), which assesses students worldwide as they reach the end of compulsory schooling. Much of the research in this area, however, draws on samples from America, and all studies cited herein are U.S.-based unless otherwise noted.

A consensus statement issued by major researchers in the area of gender-related cognitive differences identified research into the sources of individual differences in STEM achievement as an important priority (Halpern, et al., 2007). When men and women are compared at the population level, reviews find no evidence of gender differences in general intelligence (Halpern & Lamay, 2000; Neisser, et al., 1996). However, researchers have frequently observed gender differences in more specific components of cognitive ability (Boyle, Neumann, Furedy, & Westbury, 2010a, 2010b; Neumann, Fitzgerald, Furedy, & Boyle, 2007; Neumann, Sturm, Boyle, & Furedy, 2010). The size of such differences ranges
from small to large, as a function of the cognitive component under investigation (Halpern, Beninger, & Straight, 2011). The largest and most consistent gender differences are found in spatial ability (Halpern, 2011; Kimura, 2000; Maccoby & Jacklin, 1974), where reviews find effect sizes ranging from medium to large (Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995). Gender differences in spatial ability are also found cross-culturally in large international studies with young-adult samples (Peters, Lehmann, Takahira, Takeuchi, & Jordan, 2006; Silverman, Choi, & Peters, 2007).

The present review explores one such contribution to the development of spatial ability, that of gender roles. This term has been previously referred to in the literature as sex roles (Bem, 1981; Constantinople, 1973), but the term gender roles is preferred as it is broader and encompasses sociocultural factors as well as biological explanations for observed differences (Frieze & Chrisler, 2011). Relationships between spatial ability, quantitative reasoning and gender roles are discussed before reviewing empirical support for gender-role associations with spatial ability. All studies cited herein are U.S.-based unless otherwise noted.

**Spatial Ability and Quantitative Skills**

Many researchers (e.g. Wai, Lubinski, & Benbow, 2009) have proposed that spatial ability provides a foundation for the development of quantitative reasoning such as science and mathematics (Nuttall, Casey, & Pezaris, 2005; Serbin, Zelkowitz, Doyle, Gold, & Wheaton, 1990). Factor analyses of cognitive ability tests show high loadings for mathematical performance against a spatial factor (Carrol, 1993; Halpern, 2000). Furthermore, measures of spatial ability have predictive validity, in that they can predict future performance in quantitative fields (Williams & Ceci, 2007). For example, Shea, Lubinski and Benbow (2001) followed a large group of intellectually talented boys and girls over a 20 year longitudinal study, from seventh grade until age 33. They found that individual
differences in spatial, verbal, and quantitative reasoning in adolescence predicted educational and vocational outcomes two decades later. Further, spatial ability made a significant unique contribution even after controlling for verbal and mathematical ability (Shea, et al., 2001). Spatial ability is also predictive of college mathematical entrance scores (Casey, Nuttall, Pezaris, & Benbow, 1995; Casey, Nuttall, & Pezaris, 1997), which are an important prerequisite for entry to further education in science and mathematics disciplines (Ceci, Williams, & Barnett, 2009).

Factors that influence spatial ability during development hold promise for educational interventions that seek to reduce the gender gap in science and mathematics in adulthood (Halpern, 2007; Newcombe, 2007). Hyde and Lindberg (2007, p. 29) argued that even mild improvement in spatial ability may have “multiplier effects in girls’ mathematical and science performance”. Additionally, higher levels of spatial ability are associated with attitudinal changes towards mathematics and self-confidence in mathematical ability from elementary school (Eccles, Wigfield, Harold, & Blumenfeld, 1993) to high school and college (Eccles, 1987; Eccles, Jacobs, & Harold, 1990). Thus the contribution of spatial ability to later cognitive development may be in part social as well as intellectual (Crawford, Chaffin, & Fitton, 1995; Nash, 1979). Academic domains where one feels competent and are seen as being socioculturally valued for one’s gender are more likely to be pursued than those that are not (Eccles, et al., 1990).

Although medium to large gender differences in spatial ability performance are found in most reviews of studies (Linn & Petersen, 1985; Voyer, et al., 1995), Hyde (2005) notes that within-gender variation is larger than between-gender differences. Since gender alone explains only a portion of individual variation in spatial ability (Caplan & Caplan, 1994), identifying other developmental factors which promote spatial ability is an important research goal (Halpern, et al., 2007; Hyde & Lindberg, 2007). Neisser et al. (1996, p. 97) argued that
understanding the source of such differences is critical, and that such questions are “socially, as well as scientifically important”. One potential source of individual differences is that of gender-role identity.

Although the exact mechanisms contributing to the emergence of gender differences in spatial ability are debated (see Caplan & Calpan, 1994 and Halpern, 2011 for a discussion) they are believed to be influenced by a network of biological and sociocultural contributions (Ceci, et al., 2009; Crawford, et al., 1995; Eagly & Wood, 1999; Halpern & Tan, 2001). One such contribution is that of gender-role identity.

Though boys and girls typically differ in early socialisation experiences (Eccles, et al., 1990; Emmott, 1985; Lytton & Romney, 1991), there is considerable individual variation in the degree to which they develop and acquire stereotypically masculine and feminine personality traits, behaviors and interests (Bem, 1974; Constantinople, 1973; Kagan, 1964a). This process is referred to as gender typing (Kohlberg, 1966; Kohlberg & Ullian, 1974), and holds implications for the development of gender-role identity and integration of masculinity and femininity into an individual’s self-concept and gender schema (Bem, 1981; Knafo, Iervolino, & Plomin, 2005; Spence, 1993). Highly gender typed individuals are motivated to keep their behavior and self-concept consistent with traditional gender norms (Bem, 1975; Bem & Lenney, 1976; Maccoby, 1990; Martin & Ruble, 2004), and this also applies to academic domains (Nosek, Banaji, & Greenwald, 2002; Oswald, 2008; Steffens & Jelenec, 2011). Others may integrate aspects of both masculine and feminine identification into their self-schema, termed androgyny (Bem, 1984; Spence, 1984).

**Sex-role mediation of Spatial Ability**

Nash (1979) proposed a sex-role mediation explanation for gender differences in which it is argued that gender-role identity can either promote or inhibit optimum development of cognitive ability in highly gender-typed domains, such as spatial and verbal
ability. Specifically, Nash (1979) theorized that masculine identification leads to cultivation of spatial, mathematical, and scientific skills, whereas feminine identification facilitates verbal and language abilities.

In a review of gender-role influences on cognitive ability, Nash (1979, p. 263) wrote “For some people, cultural myths are translated into personality beliefs which can affect cognitive functioning in gender-typed intellectual domains”. This argument was based on earlier work by Sherman (1967) into differential learning and practice experiences of boys and girls. In doing so, Nash extended Sherman’s theory by placing cognitive development of spatial ability in a social context, where gender-role identity encourages or discourages optimum development of spatial potential. Nash identified several mechanisms that contribute to spatial development, including gender typing of intellectual domains, gender-role conformity and self-efficacy beliefs.

**Differential spatial experiences.** Sherman (1967) hypothesized a causal explanation for the presence of gender differences in spatial ability, based on a child’s differential opportunities to develop and refine spatial skills through play and recreational activities. Boys and girls typically differ in their socialisation experiences, and are encouraged by parents to engage in either stereotypically masculine or feminine play appropriate to their gender (Eccles, et al., 1990; Lytton & Romney, 1991). However, play is also an opportunity for active engagement and cognitive development (Piaget, 1968). Caplan and Caplan (1994) argued that traditionally “masculine” typed activities promote the development of spatial ability by encouraging the practice and application of spatial skills (Connor & Serbin, 1977). In contrast, traditionally “feminine” activities do not require the use of spatial skills, but reinforce other socially valued skills (Lever, 1976).

What distinguishes Sherman’s (1967) explanation from other explanations (such as Caplan and Caplan, 1994) is that it focuses specifically on gender roles, rather than solely on
biological gender, as explaining *individual* differences in spatial ability. Differential practice of skills promoting spatial development occur through gender typing of activities and interests (Serbin & Connor, 1979; Serbin, et al., 1990). Rather than assuming that the lives of boys and girls do not overlap, or that *all* boys engage in a high level of activity and receive equal opportunities to practise and develop spatial ability, it accounts for individual differences and gender typing. There is evidence to support this argument. Retrospective studies have shown that an association exists between spatial ability and activity preferences in young adult college-level samples (Baenninger & Newcombe, 1989; Signorella, Jamison, & Krupa, 1989).

**Gender typing of intellectual domains.** Kagan (1964b) noted that objects in the everyday world, social activities, and even intellectual pursuits become gender typed as either masculine or feminine, based on shared consensual beliefs that emerge very early in childhood. For example, reading and language is regarded as being feminine (Dwyer, 1973, 1974), whereas mathematics, science and technology are regarded as masculine (Li, 1999; Nash, 1975). Both at an implicit (Lane, Goh, & Driver-Linn, 2012; Nosek, et al., 2009; Steffens & Jelenec, 2011) and an explicit level (Benbow, 1988; Halpern & Tan, 2001), cultural beliefs about specific cognitive tasks as being inherently *masculine* or *feminine* prevail - even for generations growing up with increased gender equality (Liben, Bigler, & Krogh, 2002). Recently, Halpern, Straight, and Stephenson (2011) showed that lay beliefs about cognitive gender differences in student and community samples were firmly entrenched across both men and women. Although these stereotypes are not an accurate reflection of reality, Nash (1979) argued they have the potential to shape the self-concepts of boys and girls, and how they see themselves in relation to these academic domains (Hyde & Lindberg, 2007).
**Gender-role conformity pressures.** Gender roles and associated stereotypes describe differences between men and women, and prescribe how they should behave in social and occupational settings (Eagly & Mitchell, 2004). Highly gender typed persons are motivated to keep their behavior consistent with internalised gender-role standards and norms (Bem & Lenney, 1976), whereas those low in gender typing or for whom gender-role identity is less salient show greater cognitive and behavioral flexibility (Arbuthnot, 1975; Bem, 1975; Stein & Bailey, 1973). Conformity cues as to who should engage in certain behaviors, and what activities are permissible for boys or girls, come from peers, parents, and the media (Martin & Ruble, 2004; Matthews, 2007), and this has implications for intellectual domains that are masculine or feminine dominated (Eccles, 2007).

Nash (1979) argued that the increased saliency of gender and gender typing of academic subjects in adolescence may lead to a conflict between the “ideal” image a student holds of himself or herself, and the activities he or she chooses to perform well in and values. Perceived incompatibility between being "feminine" and succeeding in stereotypically “masculine” domains can hinder academic achievement (L. Rosenthal, London, Levy, & Lobel, 2011; Schmader, 2002). Thus there is also an attitudinal and motivational component to development of intellectual abilities (Nash, 1979).

**Self-efficacy beliefs and gender stereotypes.** During childhood when gender-role saliency is low, boys and girls show relatively little difference in intellectual abilities, and what differences exist often favors girls (Halpern, 2000; Nash, 1979). However, gender typing of intellectual pursuits quickly emerges in adolescence (Dwyer, 1974; Kagan, 1964b), and leads to several negative psychological consequences for some children (Nash, 1979). Firstly, girls and boys receive different messages about occupational aspirations and the usefulness of specific academic skills (Fennema & Sherman, 1977; Hyde & Lindberg, 2007). Secondly, as noted earlier, gender typing of intellectual tasks is often seen as being
incompatible with a feminine gender-role identity at a time when conformity pressure increases (Eccles, 2007; Hoffman, 1972; L. Rosenthal, et al., 2011). This can result in lowered self-esteem and reduced self-efficacy beliefs for gender-typed tasks (Pajares & Miller, 1994). Gender stereotypes suggest that men and women are better at some tasks than others, and this is reflected in self-estimations of intelligence in gender-typed domains (for a review see Szymanowicz & Furnham, 2011). Additionally, a large body of research has observed that a feminine gender typing is associated with considerably lower self-esteem than masculine or androgynous individuals (Spence, Helmreich, & Stapp, 1975; Whitley, 1983, 1988), including academic self-esteem (Alpert-Gillis & Connell, 1989; Lau, 1989; Robison-Awana, Kehle, & Jenson, 1986).


**Evidence for a Spatial-Gender-Role Association**

A prior meta-analysis by Signorella and Jamison (1986) found support for Nash’s hypothesis in spatial ability. However there have been major and potentially relevant changes in gender roles and stereotypes in the intervening decades (Auster & Ohm, 2000; Hyde & Lindberg, 2007) which Feingold (1988) has argued are responsible for declining gender differences in cognitive ability. This view is supported by Hyde (2005, 2006; 2007) and colleagues across a range of intellectual abilities (Hyde, 2007a; Lindberg, Hyde, Petersen, & Linn, 2010). These changes question the validity of Nash’s theory in contemporary society and whether such gender-role associations still exist today. For this reason, we aimed to conduct a meta-analysis of studies published since Signorella and Jamison’s (1986) review, to see whether the sex-role mediation hypothesis still holds. Although these studies are primarily based on research conducted in the USA, studies from other nations (e.g., Poland, Croatia, United Kingdom, Canada) are also examined for a broader test of Nash’s theory.
Meta-analysis provides researchers with a way to critically evaluate the cumulative evidence of empirical evidence (R. Rosenthal, 1984), and the technique is becoming increasingly common in psychology (Hyde, 1990; R. Rosenthal & DiMatteo, 2001). Although individual studies taken in isolation might show that a relationship between factor $X$ on ability $Y$ may be present or absent, factors such as random sampling error and lack of statistical power may result in erroneously rejecting the null hypothesis (Type I error) or failing to detect an effect that is real (Type II error). The technique of meta-analysis allows one to draw firmer conclusions about the existence of an association (R. Rosenthal & DiMatteo, 2001), as well to arrive at an estimate of its size that is more accurate and reliable than could be determined from a single empirical study.

A requirement of meta-analysis is that empirical studies measure a similar construct drawn from similar samples (R. Rosenthal, 1984, 1995), and that there are a sufficient number of studies to make meaningful conclusions. Spatial ability is not a unitary construct; it encompasses at least three separate processes – spatial perception, visualisation, and mental rotation (Linn & Petersen, 1985). Mental rotation is one of the most widely researched areas of cognitive gender differences (Halpern & Lamay, 2000), due in part to the fact comparisons of men and women in mental rotation show the largest effect sizes of all spatial tasks (Voyer, et al., 1995). Some researchers regard mental rotation to be a representation of general spatial reasoning (Casey, et al., 1995; Halpern, 2000; Vandenberg & Kuse, 1978), and there is evidence that performance in mental rotation prospectively predicts later development of quantitative reasoning (Casey, et al., 1997; Nuttall, et al., 2005). Therefore this review is confined to studies that investigated performance in mental rotation tasks. In addition, gender differences are larger after late adolescence when gender roles become particularly salient (Nash, 1979). There are also issues of reliability and validity when assessing gender roles in younger samples. For this reason, only studies using high school,
college or young adult samples were considered for inclusion in the reported meta-analysis. Studies using younger samples, such as that by Titze, Jansen and Heil (2010), were not considered.

In sum, the present review involved a meta-analysis of studies that have investigated gender-role associations with mental rotation task performance. It was hypothesized (Hypothesis 1) that masculinity would be positively associated with greater mental rotation performance in men and women. The influence of femininity was also investigated as a research question. It was hypothesised (Hypothesis 2) that there would be a negative association between femininity and mental rotation performance for both genders. Since the magnitude of gender differences typically varies with the type and level of difficulty of mental rotation task (Voyer, et al., 1995), we also examined the type of mental rotation instrument as a potential moderator. Similarly, because there have been debates over which measures of masculinity and femininity are the best predictor of behavior (Bem, 1984; Spence & Buckner, 2000), we examined the type of gender-role instrument as a potential moderating variable.

**Method**

**Search Strategy**

To access as many studies as possible, a number of search strategies were used. Firstly, a Web of Science citation search for articles citing either Nash (1979) or Signorella and Jamison (1986) was performed, as any study published that is relevant to the meta-analysis would be likely to cite these key articles. Secondly, GoogleScholar and PsycINFO searches were performed for studies containing the keywords “spatial ability” or “mental rotation” and any combination with the keywords “masculine”, “masculinity”, “androgynous”, or “androgyny”. This second method identified a number of additional studies that were not specifically testing a sex-role mediation hypothesis, but merely included
a gender-role measure and mental rotation task as part of a larger battery of neuropsychological tests (e.g., Rahman, Wilson, & Abrahams, 2004). Furthermore, an attempt to locate unpublished studies was made by searching the Dissertation Abstracts and ERIC databases for studies, locating one additional study. The search was performed in September, 2012.

Selection Criteria

The following inclusion criteria were used:

- peer-reviewed empirical studies published after 1986 or unpublished manuscripts and reports dated after 1986
- gender-role identity was measured using a psychometrically valid and reliable gender-role instrument, such as the Bem Sex Role Inventory (BSRI; Bem, 1974) or the Personal Attributes Questionnaire (PAQ; Spence, Helmreich, & Stapp, 1974)
- participants sampled were either an adult or high school aged adolescent, from a non-clinical sample

Requests to authors (n = 5) for additional information were made where a masculinity and mental-rotation association was not explicitly tested or reported. Three studies could not be included due to insufficient information to determine an effect size (Evardone & Alexander, 2009; Tuttle & Pillard, 1991; Vonnahme, 2005). One practice sometimes adopted is to consider all studies missing an effect size to have an association with an absolute value of zero, a practice that Rosenthal (1995) considers overly conservative and leads to inaccurate estimates. This practice was considered at length by Hedges and Becker (1986) who caution against missing value substitution. Accordingly the decision was made to exclude these missing studies. Following application of the selection and exclusion criteria, there were 12 available studies examining mental rotation and gender roles. However it should be noted
that the possibility of unpublished null studies (commonly termed the “file drawer problem”) is addressed using meta-analytic techniques that test for publication bias (Orwin, 1983; R. Rosenthal, 1979).

**Sample Characteristics**

The characteristics of all studies identified in the literature search are presented in Table 1. Several of the studies recruited participants from different countries, making for a broader test of Nash’s hypothesis than would be possible if analysing only data from the USA. It should be noted that in most studies, samples were drawn almost exclusively from student subject pools, limiting generalisability somewhat to a young-adult, college-level educated sample.

**Procedure**

Comprehensive Meta Analysis (CMA) V2 software was used for the calculation of statistics (Borenstein & Rothstein, 1999). A random-effects model was chosen (Borenstein, Hedges, Higgins, & Rothstein, 2009) because spatial ability is subject to a large number of psychosocial moderators, and a variety of different gender-role instruments and mental rotation tasks were used over multiple decades. The random effects model gives slightly wider confidence intervals than a fixed-effects model (Field, 2001; R. Rosenthal & DiMatteo, 2001), but gives a more appropriate estimate of how much variability is present in empirical studies (Kelley & Kelley, 2012).

The focus of the review was the relationship between gender-role identity and mental rotation, which can be represented by Pearson’s product moment correlation, \( r \). Gender-role instruments offer separate masculinity and femininity scales, allowing us to consider the effect of masculinity independently of femininity, and to test both for a mental rotation association.
Where the direct product-moment correlation between gender-role masculinity scale and mental rotation was reported, this was used because it represents the direct association independent of a subject’s femininity scale. However, two studies reported only the mean values for masculine, feminine, and androgynous groups. Since androgyny represents a “special case”, and some theorists argue that such participants cannot be legitimately combined with either the masculine or feminine group (Taylor & Hall, 1982), the androgynous participants were excluded as per Signorella and Jamison’s (1986) recommendation. Such an approach is the most conservative strategy available, and may lead to an underestimation of the true effect size in cases where androgynous participants (high masculinity, high femininity) score higher than their masculine or feminine counterparts (e.g. Hamilton, 1995). By doing so, however, it affords a simple comparison between masculine and feminine participants only, allowing for the use of Cohen’s $d$ and then conversion to $r$ as the common effect size unit using the formula given by Rosenthal (1984). Several studies recruited male or female participants only, and in several cases examined only masculinity associations. Calculations were performed using the CMA software.

**Meta-analytic Results**

Study characteristics and effect sizes are presented in Table 1. Since empirical studies using gender roles frequently find gender × gender-role interactions, the associations with masculinity and femininity are reported separately for men and women. Forest plots are provided when gender-role associations are statistically significant. A forest plot conveys a visual representation of the effect size estimates of individual studies and their variability (Lewis & Clarke, 2001); one can see the amount of variation between individual studies as well as the overall trend. In the centre of each study’s confidence interval is a square; the size of the square corresponds to the sample size used in each study. The diamond symbol
represents the overall estimate of the sample, with the centre of the diamond being the point estimate and its horizontal tips representing the confidence interval.

**Girls and Women**

Figure 1 presents a forest plot of the association between masculinity and mental rotation performance for girls and women, and effect sizes are given in Table 1. Hypothesis 1 predicted that masculinity would be positively associated with greater mental rotation performance. As shown in Figure 1, most studies with female samples were in a direction consistent with this hypothesis with the exception of two studies: Gilger and Ho (1989) found no association, whereas Ritter (2004) found a weak negative association. The distribution of effect sizes across studies was heterogenous, $Q(10) = 21.13$, $p = .020$, $I^2 = 52.67$ indicating moderate variability across studies. It is also noteworthy that the two largest associations were found in the non-USA samples of Croatia ($r = .64$) and Canada ($r = .45$). However, the size of the correlation is unlikely to be culturally related given that the third largest association was found in a USA sample ($r = .40$) and that small associations were also found in non-USA samples (e.g., $r = -.18$ for Ritter, 2004; $r = 14$ for Hamilton, 1995).

In support of Hypothesis 1, the combined masculinity effect size for women was $r = .23$ (95% CI lower = .11, upper = .34), $Zma = 3.72$, $p < .001$. This correlation for women was only slightly larger than that found by Signorella and Jamison (1986), who found a significant association of $r = .19$ between masculinity and mental rotation for girls and women using androgyny measures. To put these findings into perspective, we employed Rosenthal’s Binomial Effect Size Difference (BESD; R. Rosenthal & Rubin, 1982), a metric that represents effect size in a format suitable for interpretation by non-statisticians (R. Rosenthal & DiMatteo, 2001). Represented in the BESD format, the likelihood of being average or higher in mental rotation performance increases from 38.5% for feminine women to 61.5% for masculine or androgynous women.
The possibility of unpublished null studies (referred to as the “file drawer problem”) was also addressed by the calculation of Orwin’s Fail-Safe $N$, which estimates the number of null studies required to reduce mean effect sizes to a specific cutoff-point (Borenstein, et al., 2009; Orwin, 1983). Employing Orwin’s calculation, it would take only two more null studies to reduce the association to that found previously by Signorella and Jamison (1986); therefore the stronger association in these studies should be taken only tentatively.

Hypothesis 2 predicted that there would be a significant negative association between femininity and mental rotation performance. This hypothesis was not supported, $r = -.05$, $p = n.s$. Such a finding is also consistent with the findings of Signorella and Jamison (1986) who failed to find any association between femininity and mental rotation performance.

**Boys and Men**

The forest plot of the association between masculinity and mental rotation performance for boys and men is shown in Figure 2 and it presents the second test of Hypothesis 1. As can be seen from the figure, the scores of men were slightly wider in variability than for women, with many studies showing relatively large associations while three showed relatively weak or non-significant correlations. Similar to the results for women, there did not appear to be a strong relationship between the country the study was conducted in and the size of the association. The largest association was found in a sample of men from Canada ($r = .50$), but the equal second largest association was found in a USA male sample ($r = .45$). However, it is noteworthy that the two remaining studies with USA samples did not find any significant association ($r = .08$ and $r = .00$). The distribution of effect sizes across all studies was heterogenous, $Q(8) = 17.92$, $p = .022$, $I^2 = 55.36$, indicating moderate variability between studies.

In support of Hypothesis 1, the association between masculinity and mental-rotation performance for men was significant, $r = .30$, (95% CI lower = .16 upper = .42), $Zma = 4.25$,
Again, the association is slightly larger than that estimated by Signorella and Jamison (1986), who reported an \( r = .15 \) between masculinity and mental rotation performance for boys and men. Orwin’s Fail-safe \( N \) showed that it would take an additional eight unpublished studies with a mean association of zero to reduce this correlation to the size found in the earlier review (\( r = .15 \)). Represented in the BESD format, the likelihood of being average or higher in mental rotation performance increases from 35\% for feminine boys and men to 65\% for those with a masculine or androgynous gender-role identity. Finally, in contrast to Hypothesis 2, no association was found between femininity and mental rotation for boys and men, \( r = -.06, p = n.s. \).

**Moderating variables**

Since there was moderate between-study heterogeneity in the masculinity association for both men and women, it is important to determine potential moderators that may be responsible such as the type of gender-role instrument used to classify participants, or the nature of the mental rotation task. Alternately, instruments might vary in their predictive validity for men and women, and this information might be useful in planning future research. Accordingly, effect sizes and heterogeneity were examined for men and women separately across gender-role instrument.

Tables 2 and 3 present associations across type of gender-role instrument for men and women respectively. While the BSRI was used most frequently, the strongest gender-role associations were found with the PAQ for both men and women. However with an insufficient number of studies employing gender-role measures other than the BSRI, any conclusions made about the predictive validity of these instruments are tentative.

Another potential source of heterogeneity is the nature of the mental rotation task employed. Meta-analytic reviews have found that the magnitude of gender differences differs across instruments (Voyer, et al., 1995). It seems likely, therefore, that similar variation
would be present when considering gender-role associations. Table 4 presents effect sizes for studies grouped by mental rotation instrument. Instruments were grouped into four categories. These groupings reduced heterogeneity, suggesting that much of the variability observed across studies was the result of using different instruments for measuring mental rotation. It should also be noted that the Vandenberg instrument also produced the highest gender-role effect size of any mental rotation task. This may reflect the increased difficulty of this instrument which allows for greater differentiation between high and low ability (Voyer, et al., 1995).

**Discussion**

The present meta-analysis examined evidence for Nash’s (1979) sex-role mediation hypothesis of spatial ability, as measured by performance on mental-rotation tasks. In a previous review, Signorella and Jamison (1986) found a small but statistically significant association between gender role and mental rotation performance. The present results support the conclusions drawn by Signorella and Jamison (1986). There is a significant and medium sized association between masculinity and mental rotation in research conducted in the past 25 years. The size of the association did not appear to be strongly related to the country in which the study was conducted, although there was some evidence that the type of mental rotation task and gender-role measure used in the study was a factor. The present meta-analysis also showed that there was no association between femininity and mental rotation performance.

The results of this meta-analysis demonstrate three important things. Firstly, it upholds the claims made by Nash (1979) that, at least for mental rotation tasks, masculine gender roles contribute to the development of spatial ability. Although only correlational in nature, the inclusion of the longitudinal study by Newcombe and Dubas (1992) shows that gender roles have predictive validity for later development of spatial ability. Secondly, this
review demonstrates the persistence of gender roles over a larger span of time, in that studies reviewed are drawn from three decades of research; it would appear that the empirical findings of Nash and others were not a statistical quirk, or an artefact of prevailing gender inequalities of the past. Thirdly, the review shows that the magnitude of the gender-role association may be somewhat larger than previously thought by researchers, especially for men.

A possible explanation for finding a stronger association between gender roles and spatial ability than Signorella and Jamison (1986) is the quality of instruments used across studies. Many of the earlier studies reviewed by Signorella and Jamison (1986) used instruments that operationalised masculinity and femininity as bipolar opposites of a unidimensional construct (Constantinople, 1973) rather than orthogonal aspects of gender-role identity (Bem, 1981). This leads to misclassification of masculine, feminine, and androgynous participants (Bem, 1974, 1977) and an attenuation of effect size due to imprecision (H. M. Cooper, 1981). It is difficult to suggest a theoretical reason why gender roles might influence cognitive development more strongly now in men than in previous decades, but this possibility cannot be ruled out entirely.

A growing trend in empirical research is a move away from levels of statistical significance towards evaluations of the magnitude of effect sizes (Wilkinson, 1999), to assess their practical impact and importance. Cohen (1988) provides a good rule of thumb to gauge associations by: correlations of .10 or higher are regarded as small, .30 or higher as medium, and correlations higher than .50 are considered large. Frequently these yardsticks are used rather rigidly, and some researchers regard differences that are “small” as “trivial” or non-existent (Hyde, 1996, 2005). Cooper (1981) warns against this practice, as the magnitude of effects that may be found can differ greatly from one field of psychological research to another. Similarly, in a review of effect sizes and practical importance for research with
children, McCartney and Rosenthal (2000) advise against such yardsticks, and caution that effect sizes should be compared to those found in that particular research domain. For this reason, comparisons to a range of other effects deemed previously to be influential in spatial ability may be better able to put the results of this meta-analysis in context (Hyde, 1990).

The present results showed a gender-role association of $r = .30$ for men and $r = .23$ for women. Two areas previously documented to contribute to spatial ability are prior spatial activity preferences in childhood (Signorella, et al., 1989) and socioeconomic status (Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). A meta-analysis by Baenninger and Newcombe (1989) produced an $r = .10$ between spatial activity preferences and spatial ability. Levine et al. (2005) found that spatial ability differences are found between low, medium, and high socioeconomic status groups for adolescents with an effect size $r = .23$ for mental rotation. When compared to these factors, which researchers have previously argued to be important and have a meaningful impact on spatial ability, the contribution of gender role and mental rotation is greater, and may go some way to explaining existing gender differences in spatial ability (Nash, 1979; Sherman, 1967).

**Implications for spatial development**

One possible intervention being considered for individuals most at risk of forestalled spatial development is that of spatial training. In a review article, Newcombe and Frick (2010) stress the importance of early intervention in the development of spatial abilities during early childhood. Although it would be desirable to offer spatial instruction and training for all students to address this gender-gap (Hyde & Lindberg, 2007), competing interests in an ever crowded curriculum make the likelihood of this practice being adopted rather bleak; indeed few schools incorporate spatial ability specifically into the curriculum during elementary school (Mathewson, 1999; Newcombe & Frick, 2010). A more practical
measure might be for limited intervention programs to target at-risk students, in the same way that reading and literacy interventions are offered for students struggling in these areas.

While screening directly for spatial deficits may be possible, large gender differences do not typically emerge until adolescence (Linn & Petersen, 1985; Voyer, et al., 1995). Early intervention is desirable before such differences emerge (Newcombe & Frick, 2010). The assessment of gender roles might serve as a more useful risk factor to consider than gender, and it has the advantage of not necessarily being limited to one gender. Nuttall, Casey, and Pezaris (2005) describe gender-role appropriate intervention programs that develop spatial expertise, but as of yet, there are no longitudinal studies of such programs. Educators may wish to be mindful to include a range of opportunities that encourage spatial development as well as stressing their importance and relevance to both boys and girls.

Newcombe and Frick (2010) also advocate early intervention by parents, in providing children with activities and opportunities outside the classroom to develop spatial awareness, perception and visualisation. Rigidly held gender roles restrict children’s self-selection of activities (Ruble, Martin, & Berenbaum, 2006; Tracy, 1987), and parents may wish to encourage a broader repertoire in their children including sports and toys that encourage spatial development (Doyle, Voyer, & Cherney, 2012). The continuing failure to find a negative relationship between femininity and spatial ability for both genders is also noteworthy. Feminine identification should not be discouraged in order to develop spatial and quantitative ability.

**Future directions for research**

Although gender differences in cognitive ability are frequently debated, many researchers note there is greater within-gender variability than between men and women (Hyde, 1990; Priess & Hyde, 2010). Gender-role identity appears to be an important, but previously underestimated contributor to these individual differences in spatial ability, which
in turn is a key foundation for higher-level quantitative skills such as mathematics (Casey, et al., 1997; Delgado & Prieto, 2004) and STEM related fields (Halpern, 2007; Newcombe, 2007). Indeed, Halpern (2007, p. 125) has claimed that spatial ability is “essential” for success in STEM-related subjects. As such, the emergence of gender roles as a factor that meets or exceeds other factors that contribute to spatial ability is important, both as a potential diagnostic indicator for interventions as well as a focus for future investigation. By better understanding the psychosocial processes associated with gender roles and intellectual development, one might be able to identify strategies - such as self-efficacy training or challenging of gender stereotypes - that would help negate performance impairments.

Additionally, this meta-analysis affirms the merit of considering gender roles, rather than just biological gender, in studies of individual differences in cognition. Though this review was confined to only mental-rotation, it remains to be seen whether the results can be generalised more widely to other spatial ability tasks such as spatial perception and visualisation (Linn & Petersen, 1985). For example, is there something specific about a masculine or androgynous gender role that leads to improved ability to perceive spatial objects and mentally rotate them, or can it be generalised to other spatial tasks? This would allow us to test whether gender-role differences in perception are chiefly responsible, or whether there are differences in the actual cognitive processes underlying such tasks, for example a general cognitive style (Arbuthnot, 1975; Milton, 1957). A limited number of studies with adolescents and young adults have considered the Piaget water-level task (Jamison & Signorella, 1980; Kalichman, 1989; Popiel & De Lisi, 1984; Signorella & Jamison, 1978) or the Embedded Figures Test (Bernard, Boyle, & Jackling, 1990; Brosnan, 1998; Hamilton, 1995), with some inconsistencies, but larger studies are required. Furthermore, as Signorella and Jamison (1986) note, Nash’s (1979) hypothesised associations
between gender-role identity and verbal ability remain largely untested, which future studies should pursue.

**Conclusion**

We have seen many changes in society’s beliefs about gender equality in the intervening decades since Nash (1979) proposed her sex-role mediation hypothesis of intellectual development. However, for spatial ability at least, this association seems as relevant today as when the claim was first made. The results from our meta-analysis support Nash’s hypothesis for the development of spatial ability, and this provides strong support for calls to conduct further research in this area to investigate the cognitive and social processes that underlie the association between gender-roles and cognitive abilities.
Figure 1. Forest plot of masculinity association with mental rotation performance for girls and women. Positive associations indicate better mental rotation performance as masculinity increases. The combined effect size is represented as a diamond shaped correlation.

Figure 2. Forest plot of masculinity association with mental rotation performance for boys and men. Positive associations indicate better mental rotation performance as masculinity increases. The combined effect size is represented as a diamond shaped correlation.
### Table 1

*Characteristics of the Studies Identified from the Literature Search on Mental Rotation Performance and Gender Roles*

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sample Type</th>
<th>Sample Age</th>
<th>N</th>
<th>Gender Role</th>
<th>Mental Rotation</th>
<th>Masculinity (r)</th>
<th>Femininity (r)</th>
<th>Masculinity (r)</th>
<th>Femininity (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamison &amp; Signorella (1987)</td>
<td>USA</td>
<td>High school students</td>
<td>8th grade</td>
<td>10 females 19 males</td>
<td>BSRI</td>
<td>CRT</td>
<td>.45</td>
<td>.04</td>
<td>.27</td>
<td>.14</td>
</tr>
<tr>
<td>Signorella, Jamison &amp; Krupa (1989)</td>
<td>USA</td>
<td>Subject pool</td>
<td>n/r</td>
<td>132 females 156 males</td>
<td>BSRI</td>
<td>CRT</td>
<td>.08</td>
<td>.01</td>
<td>.24</td>
<td>.05</td>
</tr>
<tr>
<td>Gilger &amp; Ho (1989)</td>
<td>USA</td>
<td>Subject pool</td>
<td>M=19.0</td>
<td>52 females 38 males 65 females 65 males</td>
<td>BSRI</td>
<td>TSRT</td>
<td>.00</td>
<td>-.17</td>
<td>.00</td>
<td>-.17</td>
</tr>
<tr>
<td>Voyer &amp; Bryden (1990)</td>
<td>Canada</td>
<td>Subject pool</td>
<td>M=21.0</td>
<td>65 females 65 males</td>
<td>BSRI</td>
<td>VMRT</td>
<td>.50 **</td>
<td>-</td>
<td>.21</td>
<td>-</td>
</tr>
<tr>
<td>Tuttle &amp; Pillard (1991)</td>
<td>USA</td>
<td>Community</td>
<td>Range 25-40</td>
<td>88 females 101 males</td>
<td>CPI</td>
<td>TSRT</td>
<td>n/r</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Newcombe &amp; Dubas (1992)</td>
<td>USA</td>
<td>Longitudinal</td>
<td>M=18.0</td>
<td>122 females 54 males</td>
<td>BSRI</td>
<td>SMRT</td>
<td>.12</td>
<td>-.12</td>
<td>.14</td>
<td>-.14</td>
</tr>
<tr>
<td>Hamilton (1995) b</td>
<td>United Kingdom</td>
<td>Community, school and college</td>
<td>M=18.0</td>
<td>30 males</td>
<td>BSRI</td>
<td>SMRT</td>
<td>.34 *</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Jagieka &amp; Herman-Jeglinska (1998) c</td>
<td>Poland</td>
<td>Subject pool</td>
<td>n/a</td>
<td>54 females 41 males</td>
<td>PAQ</td>
<td>VMRT</td>
<td>.45 **</td>
<td>-.02</td>
<td>.45 ***</td>
<td>-.02</td>
</tr>
<tr>
<td>Saucier, et al., (2002)</td>
<td>Canada</td>
<td>Subject pool</td>
<td>M=22.8</td>
<td>120 females 120 males</td>
<td>EPP</td>
<td>VMRT</td>
<td>.41 ***</td>
<td>-</td>
<td>.23 **</td>
<td>-</td>
</tr>
<tr>
<td>Rahman, Wilson &amp; Abrahams (2004) d e</td>
<td>United Kingdom</td>
<td>Community</td>
<td>Range 18-40</td>
<td>37 females 42 males</td>
<td>BSRI</td>
<td>SMRT</td>
<td>.34 *</td>
<td>-.26</td>
<td>-.18</td>
<td>-.14</td>
</tr>
<tr>
<td>Ritter (2004) c</td>
<td>United Kingdom</td>
<td>Subject pool</td>
<td>M=21.0</td>
<td>41 females 46 males</td>
<td>BSRI</td>
<td>CSMRT</td>
<td>-</td>
<td>-</td>
<td>.40 **</td>
<td>.00</td>
</tr>
<tr>
<td>Scarbrough &amp; Johnston (2005)</td>
<td>USA</td>
<td>Subject pool</td>
<td>M=19.6</td>
<td>41 females 46 males</td>
<td>BSRI</td>
<td>CMRT</td>
<td>n/r</td>
<td>n/r</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Group Type</td>
<td>M</td>
<td>Gender</td>
<td>Test</td>
<td>Mean</td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------</td>
<td>--------------</td>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hromatko, et al. (2008)</td>
<td>Croatia</td>
<td>Unspecified</td>
<td>24.8</td>
<td>26 females</td>
<td>BSRI</td>
<td></td>
<td>.64*** .03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evardone &amp; Alexander (2009)</td>
<td>USA</td>
<td>Subject pool</td>
<td>20.0</td>
<td>52 females 58 males</td>
<td>BSRI VMRT</td>
<td>n/r</td>
<td>n/r</td>
<td>n/r</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Calculated from p value. ** Androgynous (high masculinity, high femininity) eliminated. *** p < .001 two-tailed. 

The following abbreviations are used: CPI = California Psychological Inventory; EPP = Eysenck Personality Profile (EPP; Eysenck, Wilson, & Jackson, 1996); CRT = Card Rotation Test (French, Ekstrom, & Price, 1963); TSRT = Thurstone Spatial Relations Test (Thurstone, 1958); SMRT = Shepard and Metzler Mental Rotation Test (Shepard & Metzler, 1971); VMRT = Vandenberg Mental Rotation Test (Vandenberg & Kuse, 1978); CMRT = Cooper and Shepard MRT (L. A. Cooper & Shepard, 1973)
Table 2
Effect Size and Heterogeneity by Gender-Role Instrument for men

<table>
<thead>
<tr>
<th>Type of instrument</th>
<th>N of studies</th>
<th>Effect size (r)</th>
<th>Zma, p-value</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bem Sex-Role Inventory</td>
<td>7</td>
<td>.25</td>
<td>Z = 3.07, p = .002</td>
<td>Q(6) = 11.73, p = n.s.</td>
</tr>
<tr>
<td>Personal Attributes Questionnaire</td>
<td>1</td>
<td>.45</td>
<td>Z = 2.20, p = .028</td>
<td>N/A</td>
</tr>
<tr>
<td>Eysenck Personality Profiler</td>
<td>1</td>
<td>.41</td>
<td>Z = 2.50, p = .013</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Total heterogeneity within-groups, Q (6) = 11.74, p = .068; between-groups, Q(2) = 6.18, p = .045

Table 3
Effect Size and Heterogeneity by Gender-Role Instrument for women

<table>
<thead>
<tr>
<th>Type of instrument</th>
<th>N of studies</th>
<th>Effect size (r)</th>
<th>Zma, p-value</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bem Sex-Role Inventory</td>
<td>8</td>
<td>.21</td>
<td>Z = 2.43, p = .015</td>
<td>Q(7) = 17.07, p = .017</td>
</tr>
<tr>
<td>Personal Attributes Questionnaire</td>
<td>2</td>
<td>.30</td>
<td>Z = 1.97, p = .049</td>
<td>Q(1) = 3.02, p = n.s.</td>
</tr>
<tr>
<td>Eysenck Personality Profiler</td>
<td>1</td>
<td>.23</td>
<td>Z = 1.17, p = n.s.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Total heterogeneity within-groups, Q (8) = 20.09, p = .010; between-groups, Q(2) = 1.04, p = n.s.

Table 4
Effect Size and Heterogeneity by Mental Rotation Instrument

<table>
<thead>
<tr>
<th>Type of instrument</th>
<th>N of studies</th>
<th>Effect size (r)</th>
<th>Zma, p-value</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card Rotations Task (French, et al., 1963)</td>
<td>2</td>
<td>.22</td>
<td>Z = 1.81, p = .071</td>
<td>Q(1) = 1.44, p = n.s.</td>
</tr>
<tr>
<td>Thurstone Spatial Relations (Thurstone, 1958)</td>
<td>3</td>
<td>.21</td>
<td>Z = 1.83, p = .058</td>
<td>Q(2) = 10.38, p = .006</td>
</tr>
<tr>
<td>Vandenberg MRT (Vandenberg &amp; Kuse, 1978)</td>
<td>3</td>
<td>.38</td>
<td>Z = 4.29, p &lt; .001</td>
<td>Q(2) = 1.41, p = n.s.</td>
</tr>
<tr>
<td>Generic Mental Rotation Tasks</td>
<td>4</td>
<td>.22</td>
<td>Z = 2.43, p = .015</td>
<td>Q(3) = 3.91, p = n.s.</td>
</tr>
</tbody>
</table>

Total heterogeneity within-groups, Q (8) = 17.14, p = .029; between-groups, Q(3) = 10.47, p = .015
References


talented preadolescents: Their nature, effects, and possible causes. *Behavioral and
Brain Sciences, 11*, 169-232. doi: 10.1017/S0140525X00049670

*Personality and Individual Differences, 11*, 213-217. doi: 10.1016/0191-
8869(90)90234-I

meta-analysis*. West Sussex. UK: John Wiley & Sons, Ltd.


methods of differential and experimental psychology to study sex differences in human
cognitive psychological functions. *Perceptual and Motor Skills, 110*, 392-410. doi:
10.2466/pms.110.2.396-410

verbal and visual-spatial tasks under different hemispheric visual-field presentation
conditions. *Perceptual and Motor Skills, 110*, 396-410. doi: 10.2466/pms.110.2.396-
410

Brosnan, M. J. (1998). The implications for academic attainment of perceived gender-
appropriateness upon spatial task performance. *British Journal of Educational


Cambridge University Press.


differences in children's self- and task perceptions during elementary school. Child
Development, 64, 830-847. doi: 10.2307/1131221

doi: 10.1037/a0018053

International Journal of Women's Studies, 8, 449-456.

(2D:4D). Archives of Sexual Behavior, 38, 442-455. doi: 10.1007/s10508-007-9260-6

UK: PSI Press.

43, 95-103. doi: 10.1037/0003-066X.43.2.95

Fennema, E., & Sherman, J. A. (1977). Sex-related differences in mathematics achievement,
spatial visualization and affective factors. American Educational Research Journal,
14, 51-71. doi: 10.3102/00028312014001051

fixed-and random-effects methods. Psychological Methods, 6, 161-180. doi:
10.1037//1082-989X.6.2.161


Frieze, I. H., & Chrisler, J. C. (2011). Editorial policy on the use of the terms “Sex” and
“Gender”. Sex Roles, 64, 789-790. doi: 10.1007/s11199-011-9988-2


