Journal of Interdisciplinary Feminist Thought

Volume 2
Issue 1 *Women and Science*

Article 5

June 2007

Picking Barbie™'s Brain: Inherent Sex Differences in Scientific Ability?

Alison Nash
Department of Psychology at the State University of New York at New Paltz, New York, nasha@newpaltz.edu

Giordana Grossi
Department of Psychology at the State University of New York at New Paltz, New York., grossig@newpaltz.edu

Picking Barbie's ™ Brain: Inherent Sex Differences in Scientific Ability?

*Alison Nash & Giordana Grossi**

State University of New York at New Paltz

Contact authors: Giordana Grossi & Alison Nash

Department of Psychology

State University of New York at New Paltz

New Paltz, NY 12561

Acknowledgments. We wish to thank Amy Kesselman, Heather Hewitt, Elaine Kolitch, and Constance Schoenberg for their invaluable insights and comments on earlier versions of this manuscript.

*The order of authorship was determined randomly, as both authors made equal contributions.

INTRODUCTION

> "There was an enormous body of masculine opinion to the effect that nothing could be expected of women intellectually. Even if her father did not read out loud these opinions, any girl could read them for herself; and the reading, even in the nineteenth century, must have lowered her vitality, and told profoundly upon her work. There would always have been that assertion—you cannot do this, you are incapable of doing that—to protest against, to overcome."
>
> Virginia Woolf (*A Room of One's Own*)

While the idea that women are not endowed with men's logical and rational skills has been a recurrent refrain in the history of science (Gould, 1981), the claim that there exist inherent sex[1] differences in mathematical and scientific abilities has recently re-emerged in the field of psychology (e.g., Baron-Cohen, 2003; Geary, 1996) and in the popular culture (a famous example is the 1994 Mattel's version of Barbie™ who proclaimed "Math is hard"). Indeed, this claim has been the object of one of the most heated scholarly debates of the last forty years (see Gallagher & Kaufman, 2005, for a recent review). Last year, it was brought back to the attention of the general public by Harvard president Larry Summers, who discussed the underrepresentation of women in science and engineering in terms of sex differences in "intrinsic aptitude" for hard sciences, with socialization and continuing discrimination considered "lesser factors" (Summers, 2005).

Summers' comments prompted a large number of magazine articles, editorials, and public debates. In these, much attention was paid to scientific studies purporting to have demonstrated biological sex differences in scientific aptitude. Catch phrases like 'brain hardwiring' entered public discourse. The popular media, however, often misrepresent complex scientific findings, and furthermore, readily accept these findings at face value. Feminist scholars have long recognized the cloak of legitimacy that science drapes over the status quo (Bleier, 1991; Millet, 1970; Weisstein,

[1] We consider gender a social construct. However, throughout this paper, we use the term 'sex' rather than 'gender' as we are addressing biological theories about differences in females' and males' abilities.

1970; 1997). The alleged scientific discovery that particular areas of women's and men's brains are 'hardwired' differently serves to reify and consolidate the belief in women's natural limitations in scientific competence.

Our own training and research in the fields of cognitive neuroscience and developmental psychology enable us to critically evaluate the studies on which this claim is based. We therefore provide an overview of the research on biological sex differences in scientific ability. We show that findings of differences in math and science performance are not reliable and depend on the measures used, and furthermore, that the key evidence for biological predispositions comes from poorly designed studies with equivocal findings. Indeed, the scientific establishment itself distorts the research findings.

The argument that women's underrepresentation in science stems from inherent sex differences in aptitude rests on two claims: 1) There are measurable sex differences in scientific/mathematical aptitude (*Psychometric Claim*), and 2) these differences stem from underlying biological predispositions (*Biological Difference Claim*). We argue that these claims are not justified either theoretically or empirically. They are based on assumptions that are never fully tested or are contradicted by empirical results. We analyze each claim separately.

PSYCHOMETRIC CLAIM

Mathematical and scientific abilities have been assessed with a range of measures, including standardized tests, cognitive tasks, and performance in school. The central measure of advanced mathematical skills in the United States is the mathematical scale of the SAT (SAT-M). Halpern, Wai, and Saw (2005) note that males outperform females on the SAT-M by about 40 points and the difference in scores has remained stable over the last 25 years. This supposedly intractable gap provides key evidence for claims of inherent sex differences in abilities. But is the gap universal?

International assessments have provided valuable data on whether the sex difference observed in the United States is present in other countries. The Programme for International Student

Assessment (PISA) assesses literacy, mathematics, and problem-solving in 15-year-olds every three years (Organisation for Economic Cooperation and Development, OECD, 2004). In its 2003 assessment of 250,000 students from 41 countries, males scored higher than females in mathematical tests in 27 countries out of 40. In one country, Iceland, girls outperformed boys; indeed, girls in Iceland even outperformed boys in the U.S. Furthermore, the results for problem-solving tests indicated very few sex differences and, when present, they favored girls.

Therefore, it appears that a male advantage in mathematics and problem-solving is not universal and varies greatly across countries. Based on these findings, the report concludes that "The wide variation in gender gaps among countries suggests that the current differences are not the inevitable outcomes of differences between young males and females and that effective policies and practices can overcome what were long taken to be the inevitable outcomes of differences between males and females in interests, learning styles and, even, in *underlying capacities* [italics added]" (OECD, 2004, p. 97).

Furthermore, in contrast to standardized measures, females tend to have better grades than males in all subjects, including advanced mathematical courses (Kimball, 1989). Indeed, research has shown that the SAT-M scores underpredict the performance of females relative to males in college mathematical courses, both introductory and advanced (Chipman, 2005). For example, Wainer & Steinberg (1992) found that females who received the same grades as males in the same university math courses had received scores nearly 50 points lower on the SAT exam.

Thus, the data demonstrate that a male advantage in mathematics is not a solid finding, with different measures yielding different outcomes. Therefore, it is not clear that sex differences in standardized test scores indicate differences in aptitude. Indeed, the creators of the SAT exam themselves are mystified about exactly what the SAT measures. When it first debuted in 1901, SAT stood for Scholastic *Achievement* Test. In 1941, it was renamed by the College Board as the Scholastic *Aptitude* Test. In 1990, when "faced with the fact that a test that can be very successfully prepared for is not a true test of ability" (PrepMe© 2001-2005, History of SAT), the SAT was renamed the Scholastic *Assessment* Test. Finally, in 1994, since 'assessment test' is redundant, the test became simply the SAT. The College Board explained "Please note that SAT is not an initialism. It does not stand for anything" (PrepMe© 2001-2005, History of SAT). Given that it is not clear, even to the developers themselves, what the 'A' stands for, what can be said about what the test measures?

Furthermore, because sex differences on standardized math tests first emerge in late adolescence (see Spelke, 2005, for a recent review), these differences can reflect the influence of a variety of experiential factors, such as academic curricula, stereotype threat, and socialization:

Academic Curricula

In many countries, boys and girls attend different secondary schools and have diversified curricula from the age of 14. Because boys tend to enroll in scientific courses more frequently than girls, they take more advanced math classes than girls, which could enhance their performance on standardized tests. Similarly, in the U.S., many more boys than girls take AP classes in calculus, chemistry, physics, and computer science (U.S. Census Bureau, 2004-2005), which could explain the sex differences in SAT-M scores. It is noteworthy that in Iceland, where girls outperform boys in math and problem-solving, boys and girls take the same classes until the age of 16 (Ministry of Education, Science, and Culture in Island, 2005).

Stereotype Threat

Performance on standardized tests is sensitive to a psychological pressure termed "stereotype threat" (Aronson, Lustina, Good, & Keough, 1999), in which members of a stereotyped group (e.g., African-Americans, the elderly, and women) underperform in situations in which they are aware that their behavior can confirm their negative reputation (Inzlicht & Ben-Zeev, 2000; Levy, 1996; Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995). Specifically, if women are even subtly reminded that math is hard, as Barbie™ declared, they tend to perform poorly on math tests. Importantly, sex differences disappear when the stereotype threat is removed or minimized. For example, sex differences in performance on the quantitative GRE were eliminated when students were informed beforehand that women and men performed equally well on the test (Spencer et al., 1999). Using the same test, Inzlicht & Ben-Zeev (2000) showed that women experienced performance deficits in a test-taking environment in which men outnumbered them, and that performance improved linearly when the number of males tested in the same room decreased. By simply testing students in a room with more girls than boys, girls perform better – the fewer the boys, the better girls do. These data show that test performance is influenced by social factors that are unrelated to the ability that is being measured.

Socialization

Finally, sex differences in standardized test scores do not appear until children have experienced years of socialization that may have affected performance. There are well-documented sex differences in socialization – i.e., differential expectations, treatment, and environments for girls and boys**.** Beginning at birth, parents' perceptions of daughters and sons differ. Despite similarities in physical size, daughters are described as dainty and soft whereas boys are described as sturdy and tough (Rubin, Provenzano, & Luria, 1974). Parents (e.g., Witt, 1997), peers (e.g., Maccoby, 1990), and teachers (e.g., Dickman, 1993) continue to treat girls and boys differently. Additionally, very different environments are provided for girls and boys. Studies spanning many years (Nash & Krawcyzk, 1994; Pomerleau, Bolduc, Malcuit, & Cossette, 1990; Rheingold & Cook, 1975) indicate that girls play with toys representing the home, such as dolls and kitchen appliances, while boys' toys represent the wider world, such as transportation vehicles and construction materials (which enhance spatial skills; Baenninger & Newcombe, 1989). Similarly, children's media depict different roles for girls and boys, with male characters still dominating and in active roles, and the substantially fewer female characters in subordinate roles (Nash & Burckle, 1997; Thompson & Zerbinos, 1995). It is thus clear that cultural expectations about gender are communicated to children from birth, and that these differential expectations and experiences can affect children's interest and abilities.

In summary, it appears that findings of sex differences in scientific abilities are not as robust, reliable, and universal as they are often portrayed, and when they exist, may be explained by socialization and contextual factors. Next, we examine research that attempts to eliminate nurture from the equation, and claims to have discovered the presence of inherent sex differences before socialization begins.

BIOLOGICAL DIFFERENCE CLAIM

Support for this claim is provided by studies of newborns and studies on the effects of prenatal testosterone on brain organization. We review each of these areas of research in turn.

Newborns

As we have discussed, sex differences in math and science abilities can be due to biological and/or experiential factors: nature and nurture are always confounded. By studying very young children, experiential factors can be reduced or eliminated. Because studies indicate that even during infancy, male and female infants receive differential treatment (reviewed by Golombok & Hines, 2002), the only way to identify biological sex differences is by eliminating experience entirely and focusing on newborns. There is only one study that examines sex differences in relevant behaviors in newborns (Connellan, Baron-Cohen, Wheelright, Batki, & Ahluwalia, 2000), and this study has become the cornerstone of the argument for innate predispositions. According to Steven Pinker, a leading scholar on cognition and language, "Simon Baron-Cohen has given us some of the most sophisticated research on the nature and origin of sex differences in cognition" (Pinker, 2005). We therefore focus on this study, carefully examining its theoretical rationale, methodology, and conclusions in detail.

Theoretical Rationale

The theoretical rationale stems from Simon Baron-Cohen's work on autism (Baron-Cohen, Knickmeyer, & Belmonte, 2005), a condition characterized by social impairment and heightened interest in the physical world. Baron-Cohen views autism, more common in males than in females, as a manifestation of an “extreme male brain”. He further posits that, in general, male cognition reflects an emphasis on analysis and mechanical understanding – what Baron-Cohen terms ‘systemizing.’ Systemizing abilities, in turn, would provide the basis for scientific reasoning. In contrast, female cognition reflects an emphasis on social understanding -- what Baron-Cohen terms ‘empathizing.’ According to Baron-Cohen (2003; Baron-Cohen et al., 2005), these sex differences in ‘systemizing’ and ‘empathizing’ capacities stem from prenatal ‘hardwiring’ in the brain and are present at birth.

Strong support for this theory would be provided by demonstrating that sex differences in these capacities do indeed begin at birth. Unfortunately, although newborns provide an opportunity to examine biological predispositions prior to experience, they are in fact very difficult to study. They drift among various states of consciousness in an unpredictable manner, going from an alert state to crying to sleeping within a very short time. Their attention spans are extremely variable (Fogel, 2001). To address some of these difficulties, standard methodologies are typically used. However, despite an existing body of research on newborn perceptual preferences, Connellan and Baron-Cohen’s newborn study does not use this methodology, or address the findings from this body of research. We next address these concerns in an in-depth critique of Connellan et al.’s (2000) study.

Methodology

Connellan et al. assume that sex differences in infants’ interest in ‘social’ versus ‘mechanical’ stimuli are precursors to future sex differences in empathizing and systemizing abilities. They therefore compared one-day-old girls’ and boys’ interest in ‘social’ versus ‘mechanical’ stimuli. The social stimulus was the real, live, face of the first author of the study, Jennifer Connellan. The mechanical stimulus was a mobile – a face-sized ball composed of various facial features that were haphazardly arranged. The stimuli were presented sequentially to infants who were tested at their mothers’ bedsides *or* in the neonatal nursery, lying on their backs in their cribs *or* held in their parents’ laps. Differences in time spent looking at the face or the mobile were considered indices of preferences.

This study is fraught with methodological problems:

1) Validity: The investigators assumed that one-day-old newborns' preferences for faces or mobiles reflect social versus mechanical intelligence. However, it is not clear that looking at a face or a mobile does, in fact, reflect later social or mechanical abilities. Furthermore, even in newborns, it is unclear that a face represents a 'social' stimulus. Recent studies suggest that newborn preferences for face patterns compared to other patterns represent a perceptual bias for top-heavy patterns, rather than preferences for faces per se (Macchi-Cassia, Turati, & Simion, 2004). Indeed, other studies indicate that newborns actually prefer looking at these kinds of patterns to looking at real faces (Simion, Macchi-Cassia, Turati, & Valenza, 2001). It is not until 3 months of age that infants prefer real faces to face-like (top heavy) patterns (Turati, Valenza, Leo, & Simion, 2005).

2) Confounds/lack of control: The face and mobile differed on several crucial variables. The face in fact consists of several properties. It consists of movement and expressions, and is attached to a live person who exudes warmth and odors. As all of these vary together, it is impossible to know which of these dimensions underlie any preferences for the 'face.' Again, it is not clear the preferences for the face can be attributed to its 'social' dimension. Furthermore, infants were tested in different settings. Slight changes in the settings could lead to differences in the perception of the two stimuli, making them appear more or less top heavy under different conditions. Preferences for one stimulus or the other could simply reflect these different perspectives.

3) Experimenter expectancies: A striking design flaw is that the face stimulus was that of the researcher herself. Experimenter and subject expectancies are well-documented, and require stringent controls. In this case, the researcher could unconsciously move her face or tilt her head in ways that increase its salience, which is particular problematic given that in many cases she was aware of the newborn's sex (Edge, 2005, Simon Baron-Cohen).

4) Operational definition of the dependent variable: Curiously, the dependent variable itself was incorrectly defined. The authors report that "looking time was calculated as a proportion of *total* *looking time*" (p. 115). However, examination of the results indicates that looking time to each stimulus was actually compared to the *presentation time* for each. In order to interpret findings, the definition of the dependent variable must be clear and precise.

5) *Statistics.* Finally, we noticed a serious mistake in the number of degrees of freedom used in the statistical analysis that compared girls' looking time to the face and mobile. It is very possible that the correct number would not have yielded significant results. Precision in statistical analyses is a fundamental component of scientific investigation. We wonder why this obvious error was overlooked in the peer review process for journal publication.

These are all serious methodological problems. There are numerous investigations that adopt a standard research paradigm for investigating perceptual preferences: all infants are tested in the same, controlled setting, stimuli are presented *simultaneously* to control for the fragility in newborn attention span and fluidity of newborn states, and specific procedures are used to prevent parents and researchers from influencing infants' behavior. In studies of sex differences, experimenters are blind to the sex of the infants. Finally, the dependent variable is measured more precisely – in terms of actual looking time (the difference in the amount of time that infants looked at one stimulus compared to another) rather than relative looking time. It is unclear why the only study of sex differences in perceptual preferences in newborns ignored these standard procedures.

Findings and Conclusion

The only significant findings were that boys looked at the mobile more than girls did, and girls looked at the face more than they looked at the mobile. Much was made of these findings in the paper's conclusion: the authors state that they "demonstrate *beyond a reasonable doubt* [italics added] that these [sex] differences are in part biological in origin." (p. 114). This is a strong claim, with serious implications. The methodological problems already provide reasonable doubt. We next show that the findings themselves are equivocal in regard to this claim.

The authors conclude that "We have demonstrated that at 1 day old, human neonates demonstrate sexual dimorphism in both social and mechanical perception. Male infants show a stronger interest in mechanical objects, while female infants show a stronger interest in the face" (Connellan et al., 2000, p. 116). However, this conclusion incorporates comparisons that were *not* significant: Boys did *not* look at the mobile more than the face; they only looked at the mobile more than girls did, and girls did *not* look at the face more than boys did; they only looked at the face more than the mobile. The title of the paper, "Sex Differences in Human Neonatal Social Perception," is in fact misleading, as the findings indicated no sex difference in looking time to the face.

Nevertheless, these findings are misinterpreted and misstated in many different papers as evidence for newborn sex differences in social and mechanical preferences. For example, in his 2005 paper in the prestigious journal *Science*, Baron-Cohen asserts that "biological contributions to social interest are suggested by studies of human infants. When 1-day-old babies are presented with either a live face or a mechanical mobile, girls spend more time looking at the face, whereas boys prefer the mechanical object" (p. 820). In other papers, Baron-Cohen *incorrectly* summarizes the newborn findings stating that "among one day old babies, *girls look longer at faces than boys do* (Baron-Cohen, 2004, The Female Brain; Empathizing), and in another paper, "At birth, boys look longer than girls at a mobile whilst *girls look longer than boys at a face"* (italics added, Lutchmaya & Baron-Cohen, 2002, p. 319). Recall that neither of the highlighted findings was significant. Thus, claims of newborn sex differences in widely-read papers were supported by findings that were incorrectly reported.

This widely-cited study provides the centerpiece for the claim that sex differences in social and scientific capacities stem from biological predispositions. In addition to all its problems, in the countless studies of preferences for face-like patterns in very young infants, the Connellan/Baron-Cohen findings have not been replicated. No other study has found sex differences in preferences for face-like patterns compared to other patterns. When preferences for real faces over face-like patterns emerge in three-month-olds, they are found for girls and boys alike (Turati et al., 2005).

Baron-Cohen claimed that a second study of one-year-olds provides further support for sex differences in interest in social versus mechanical stimuli (Lutchmaya & Baron-Cohen, 2002). Although there were methodological and measurement problems in this study as well, here we will focus on one of their findings: boys preferred 'non-social' to 'social' stimuli, but girls had no preference (p. 321). These findings are not consistent with the newborn findings that girls preferred social to non-social stimuli, whereas boys had no preference.

Taken together, the lack of replication and robustness of the findings presents serious problems for the argument that sex differences in newborn social and mechanical interests are due to "hardwiring" in the brain (Baron-Cohen, 2003). Furthermore, if these sex differences reflect underlying brain organization, girls and boys paths should continue to diverge as these capacities become more sophisticated. If girls are 'hardwired' for social intelligence (empathizing), and boys for scientific reasoning (systemizing), sex differences should be found in children's emerging skills in these areas. This is not the case. There is a large body of research that charts the development of young children's emerging understanding of the physical and social worlds. A brief review of the studies of the development of systemizing and empathizing in early life gives no indication that the newborn sex differences are maintained.

The Early Development of Systemizing and Empathizing

The fields in which girls are underrepresented – science and math – are based on a set of fundamental reasoning skills: an understanding of the properties of objects, space, and quantity. If sex differences in these areas reflect differential brain organization present at birth, we would expect that the purported 'hardwiring' present in newborns would to lead to sex differences in the development of these abilities. A large body of research in this area indicates that this is not the case.

A series of well-controlled experiments by Renee Baillargeon (2004; 2005) charts the development of infants' surprisingly sophisticated knowledge of various properties of objects and mechanical motion. Interestingly, in dozens of studies of two- to twelve-month-old infants, Baillargeon found only two sex differences, both favoring *girls*: first, girls understand a month earlier than boys do that for one object to support another, they must be in contact. Second, girls understand principles of collision a month before boys do. Similarly, studies of the development of numerical (Wynn, 2002) and spatial reasoning (Newcombe, 2002) over the course of infancy reported no sex differences. There is therefore no evidence for 'hardwired' brain differences that predispose males more than females to understand physical/mechanical phenomena.

According to Baron-Cohen's theory (2003), an inherent female disadvantage in 'systemizing' is coupled with a superiority in 'empathizing', manifested in newborn girls' preference for faces. Presumably, the development of empathy itself should reflect female superiority. Studies of the emergence of empathy during the second year of life provide inconsistent findings (e.g., Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). When sex differences are found, they are only for maternal report measures; direct observations yield no sex differences (Zahn-Waxler et al., 1992). It is widely known that parental report measures are susceptible to observer bias.

Empathy becomes more fully developed once children recognize that people's understanding of the world is not based on a uniform reality, but on perceptions, beliefs, and desires. This understanding has been called a 'theory of mind,' which emerges between three and four years of age. Its absence is a central characteristics of autism, which Baron-Cohen attributed to an 'extreme male brain.' If extreme male brains lack a theory of mind, we would expect that average male brains would lag behind female brains in the development of a theory of mind. This is not the case. As Alison Gopnik, a leader and prominent researcher in the study of theory of mind, remarked, "In hundreds of studies of children's Theory of Mind there is no evidence of sex differences or a female advantage" (Edge, 2004, Alison Gopnik; for a review of theory of mind research see Wellman, Cross, & Watson, 2001).

In summary, a review of the research in 'systemizing' and 'empathizing' capacities in early life indicate similar developmental trajectories for girls and boys, providing more than reasonable doubt about the existence of biological sex differences in these areas. Thus, the developmental studies do not provide solid evidence for innate sex differences in scientific abilities. We next turn to studies of the proposed mechanism for the development of sex differences in brain organization. These studies take us further back in development, prior to birth, to the prenatal period.

Prenatal Testosterone and Brain Organization

Mathematics and scientific reasoning rely on a multitude of complex cognitive processes, including language, memory, visuo-spatial skills, reasoning and analytic skills. In turn, these processes are supported by several neural networks, which function in a highly integrated fashion. In order to claim that there exist sex differences in "intrinsic aptitude" for hard sciences, a researcher should pinpoint the specific cognitive process involved, the physiological mechanisms that underlie them, and explain how such mechanisms are responsible for differences in cognition and performance.

Several researchers (e.g., Baron-Cohen, 2003; Benbow & Stanley, 1980; Casey, Nuttall, Pezaris, & Benbow, 1995; Geary, 1996) explain sex differences in mathematics in terms of sex differences in visuo-spatial abilities. Two pieces of evidence are provided: first, males outperform females in some spatial tasks, specifically those requiring the mental rotation of three-dimensional objects (Geary, 1996; Linn & Petersen, 1995); second, spatial skills predict SAT-M scores after controlling for the SAT-Verbal scores (Casey et al., 1995).

Researchers claiming the existence of a male superiority in spatial skills have found Geschwind and colleagues' model (Geschwind & Behan, 1982; Geschwind & Galaburda, 1987) a useful framework for justifying their view that such superiority relies on a biological predisposition present at birth. The model posits that high levels of prenatal testosterone, usually present in males, slow down the growth of the left hemisphere and enhance the development of the right hemisphere in terms of both size and functions (Bryden, McManus, & Bulman-Fleming, 1994)[2]. Therefore, the right hemisphere should be larger in males than females, and the difference in size should be related to differences in visuo-spatial skills. Similarly, levels of prenatal testosterone should be related to visuo-spatial skills later in life. Visuo-spatial abilities should then predict mathematical abilities.

The evidence does not, in fact, support these predictions. In one recent study, Witelson, Beresh, and Kigal (2005) correlated Wechsler Adult Intelligence Scale (WAIS) scores with postmortem cerebral volume in 100 cases. No differences in the volume of the two hemispheres were found within or between sexes (after hemispheric measures were corrected for total brain volume), and no sex differences were found in visuo-spatial abilities (measured by the Performance Scale Score of the WAIS). Furthermore, visuo-spatial abilities were not correlated with hemisphere-related measures or differences. In a nutshell, apart from a well-corroborated difference in total brain size (men have larger brains by ~10%), no differences between the hemispheres were associated with sex or visuo-spatial abilities.

Similarly, studies focusing on various indices of brain structure in living and healthy individuals have not provided substantial evidence for hemispheric differences related to sex, not even for areas important to visuo-spatial processing (e.g., Allen, Damasio, Grabowski, Bruss, & Zhang, 2003; Barta & Dazzan, 2003). When such differences are present, their functional meaning remains unclear. Most important, because brain organization changes with experience (e.g., Maguire, et al., 2000; Recanzone, Schreiner, & Merzenich, 1993), sex differences in brain structure in adults could reflect differences in experience rather than differences present at birth.

Studies on prenatal testosterone have offered a similarly confused picture. Grimshaw, Sitarenios, and Finegan (1995) measured spatial abilities in 7-year-old children with a version of the mental rotation test. Prenatal testosterone levels had previously been measured in the amniotic fluid

[2] Please note that, according to the model (Geschwind & Galaburda, 1987), extreme low and high levels of prenatal testosterone have negative effects on the development of the right hemisphere and, therefore, of visuo-spatial abilities.

obtained by amniocentesis during the second trimester of pregnancy. Although a positive correlation was found between measures of mental rotation and prenatal testosterone levels, it was found for girls (girls with higher levels of prenatal testosterone rotated faster and had shorter reaction times than girls with lower levels), but not for boys. Importantly, no sex differences were found in rotation performance. In another study conducted with 6-year-olds, girls with *higher* levels of testosterone taken from the umbical cord at birth tended to have *poorer* spatial abilities, whereas no correlations were observed for boys (Jacklin, Wilcox, and Maccoby, 1988). Therefore, these two studies provide conflicting results.

In summary, these findings do not support the view that sex differences in visuo-spatial abilities are based on differences in brain asymmetry. The physiological mechanisms hypothesized to underlie the purported male advantage in visuo-spatial skills have not been confirmed. If structural differences related to visuo-spatial abilities exist at the microstructural level, their existence has yet to be demonstrated in humans. Moreover, there is no convincing evidence that prenatal testosterone is related to visuo-spatial functions later in life. Effects of prenatal testosterone on sexual dimorphism in brain structure and behavior have been found in the rat (e.g., Lewis & Diamond, 1995), but not in humans. Nonetheless, from this handful of studies, some researchers have concluded that visuo-spatial skills in males are enhanced compared to females due to the effects of prenatal testosterone. Baron-Cohen (2003) even misrepresents Grimshaw et al.'s (1995) results, stating that "…the higher a child's prenatal testosterone, the better they performed the Mental Rotation Test." (p. 102). Boys, on average, had higher levels of prenatal testosterone than girls, but their mental rotation scores did not differ from girls' scores. Thus, no convincing evidence exists for testosterone-related sex differences in brain structure or spatial performance. The claim that sex differences in mathematics and scientific skills stem from biological sex differences in visuo-spatial abilities remains unsupported.

CONCLUSIONS

Does women's underrepresentation in scientific fields stems from inherent sex differences in abilities? To answer this question, we examined the evidence for biological sex differences. Although the consistent sex difference in SAT-M is considered compelling, we have shown that it is unclear what the SAT-M measures, that other measures of math and scientific abilities do not favor boys, and that some measures favor girls. Furthermore, even when measures favor boys, differences in scores may reflect differences in school curricula, cultural expectations, and experiences. The internalization of cultural stereotypes clearly affects test scores, such that the removal of stereotype pressure by simply changing instructions and seating arrangements can improve girls' test performance.

Furthermore, the biological evidence for inherent sex differences in science and math ability comes from only a few studies conducted by researchers who are inexperienced with the methodology and issues in the fields of developmental psychology and cognitive neuroscience. Evidence for sex differences in newborn preferences for mechanical versus social stimuli is suspect, and claims that they are hard-wired are questionable as they disappear during the course of infancy. Furthermore, inconsistent and contradictory findings emerge from studies of the links between prenatal testosterone, brain organization, and spatial skills. Current neuroscience research emphasizes the plasticity of the brain, with neural connections that are continually changing through experience.

Despite the lack of rigor in research on biological sex differences in scientific abilities, the studies are published in prestigious journals and receive much attention in the popular media. The widespread acceptance of their conclusions may be more a reflection of biases and beliefs than of scientifically-based evidence. As scientists, we expect the same rigorous standards to be applied to studies of sex differences as other areas in the behavioral sciences. As feminists, we are deeply concerned that sloppy science is tolerated when it confirms prevailing hierarchies and cultural stereotypes. We would like to see the sophisticated tools of the behavioral sciences applied to investigations that assist all children in better understanding the scientific and social worlds.

References

Allen, J. S., Damasio, H., Grabowski, T. J., Bruss, J., & Zhang, W. (2003). Sexual dimorphism and asymmetries in the gray-white composition of the human cerebrum. *NeuroImage*, 18, 880-894.

Aronson, J., Lustina, M. J., Good, C., & Keough, K. (1999). Why white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46.

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. Sex Roles, 20 (5-6), 327-344.

Baillargeon, R. (2003). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswame (Ed.), *Blackwell Handbook of Childhood Cognitive Development* (pp. 47-64). Oxford: Oxford University Press.

Baillargeon, R. (2004). Infants' physical world. *Current directions in psychological science, 13,* 89-94.

Baron-Cohen, S. (2003). *The essential difference. The truth about the male and female brain.* New York: Basic Books.

Baron-Cohen, S. (2004, March 29). Do sex differences in empathy account for sex differences in language acquisition? *Coevolution of Language and Theory of Mind.* Retrieved from http://www.interdisciplines.org/coevolution/papers/7.

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: Implications for explaining autism. *Science, 310*, 819-823

Baron-Cohen, S., Lutchmaya, S., & Knickmeyer, R. (2004). *Prenatal testosterone in mind*. MIT Press: Cambridge, MA.

Barta, P., & Dazzan, P. (2003). Hemispheric surface area: Sex, laterality, and age effects. *Cerebral Cortex, 13*, 364-370.

Benbow, C. P. & Stanley, J. C. (1980). Sex differences in math ability: Fact or artifact?, *Science,* *210*, 1262-1264.

Bleier, R. (Ed.). (1991). *Feminist approaches to science*. New York: Teachers College Press.

Bryden, M. P., McManus, I. C., & Bulman-Fleming, M. B. (1994). Evaluating the empirical support for the Geschwind-Behan-Galaburda model of cerebral lateralization. *Brain & Cognition, 26*, 103-167.

Chipman, S. F. (2005). Research on the women and mathematics issue. A personal case history. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender Differences in Mathematics. An Integrative Approach* (pp. 1-24). NY: Cambridge University Press.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behavior and Development*, *23*, 113–118.

Dickman, C. B. (1993). Gender differences and instructional discrimination in the classroom. *Journal of Invitational Theory and Practice*, *2*, 35-42.

Dorval, M., & Pepin, M. (1986). Effect of playing a video game on measure of spatial visualization. *Perceptual Motor Skills, 62*, 159–162.

Edge (2005). *The assortative mating theory: A talk with Simon Baron-Cohen.* Retrieved from http://www.edge.org/3rd_culture/baron-cohen05/baron-cohen05_index.html.

Fogel, A. (2001). *Infancy*. Belmont, CA: Wadsworth.Gallagher, A. M., & Kaufman, J. C. (Eds.) (2005). *Gender Differences in Mathematics. An Integrative Approach*. NY: Cambridge University Press.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19***,** 229–284.

Geschwind, N., & Behan, P. (1982). Left-handedness: Association with immune disease, migraine, and developmental learning disorder. *Proceedings of the National Academy of Science, USA, 79*, 5097-5100.

Geschwind, N. & Galaburda, M. (1987). *Cerebral lateralization*. Cambridge, MA: MIT Press.

Golombok, S. & Hines, M. (2002) Sex differences in social behavior. In P.K. Smith & C.H. Hart (Eds.), Blackwell *Handbook of childhood social development* (pp. 117-136).

Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.

Grimshaw, G. M., Sitarenios, G., & Finegan, J.-A. K. (1995). Mental rotation at 7 years: Relations with prenatal testosterone levels and spatial play experiences. *Brain & Cognition, 29*, 85-100.

Halpern, D. F., Wai, J., & Saw, S. (2005). A Psychobiological model: Why females are sometimes greater than and sometimes less than males in math achievement. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender Differences in Mathematics. An Integrative Approach* (pp. 49-72). NY: Cambridge University Press.

Inzlicht, M.,& Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11* (5), 365-371.

Jacklin, C. N., Wilcox, K. T., & Maccoby, E. E. (1988). Neonatal sex-steroid hormones and cognitive abilities at six years. *Developmental Psychobiology, 21*, 567-574.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105* (2), 198-214.

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1092–1107.

Linn, M. C., & Peterson, A. C. (1985). Emergence and characterization of sex differences in spatial abilities: A meta-analysis. *Child Development, 56,* 1479-1498.

Lutchmaya, S., & Baron-Cohen, S. (2002). Human sex differences in social and non-social looking preferences, at 12 months of age. *Infant Behavior & Development, 25*, 319-325.

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin, 109*, 267-296.

Macchi Cassia, V., Turati, C., & Simion, F. (2004). Can a non specific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science*, 379–383.

Maccoby, E. E. (1990). Gender and relationships: A developmental account. American Psychologist, 45, 513-520.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., and Christopher D. Frith, C. D. (1997). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Science, 97* (8), 4398–4403.

Millet, K. (1970). *Sexual politics.* New York: Doubleday.

Ministry of Education, Science, and Culture in Island (2005). Retrieved on May 30, 2006, from http://eng.menntamalaraduneyti.is/publications/.

Nash, A., & Burckle, M. (1997). *The portrayal of gender in children's cartoons and its reflection in children's room*. Presented to the Society of Research in Child Development, Washington, DC, April.

Nash, A., & Krawczyk, R. (1994, April). *Boys' and girls' rooms revisited: The contents of boys' and girls' rooms in the 1990s.* Presented at the Conference on Human Development, Pittsburgh, PA.

Newcombe, N.S. (2002). The nativist-empiricist controversy in the context of recent research on spatial and quantitative development. *Psychological Science, 13,* 395-401.

Organization for Economic Co-operation and Development (2004). Learning for Tomorrow's World. First Results from PISA 2003. Retrieved May 31, 2001 from http://www.oecd.org/.

Pinker, S. (2005). Letter to the New York Times, *August 12, 2005*

Pomerleau, A., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first two years of life. *Sex Roles, 22*, 359-367.

PrepMe© (2001-2005). History of SAT. Retrieved April 28, 2005 from http://www.prepme.com/SAT/sat-history.htm.

Recanzone, G., Schreiner, C. E., & Merzenich, M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkey. *Journal of Neuroscience, 13*, 87-103.

Rheingold, H. L., & Cook, D. V. (1975). The content of boys' and girls' rooms as an index of parents' behavior. *Child Development*, 46, 459-463.

Rubin, J.Z., Provenzano, F.J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry, 44,* 512-519.

Simion, F., Macchi Cassia, V., Turati, C., Valenza, E., (2001). The origins of face perception: specific versus non-specific mechanisms. *Infant and Child Development*, *10*, 59 - 65

Spelke, E. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60* (9), 950-958.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Summers, L. H. (2005). Remarks at NBER conference on diversifying the science and engineering task force. Cambridge, Massachusetts, January 14, 2005. Retrieved on May 30, 2006, from http://www.president.harvard.edu/speeches/2005/nber.html.

Thompson, T. L. & Zerbinos, E. (1995). Gender roles in animated cartoons: Has the picture changed in 20 years? *Sex Roles, 32*, 651-673.

Turati, C., Valenza, E., Leo, I., Simion, F. (2005). Three-month-olds' visual preference for faces and its underlying visual processing mechanisms. *Journal of Experimental Child Psychology*, *90*, 255-273.

U.S. Bureau of the Census (2004-2005). Statistical Abstract of the United States: 2004-2005**.** Washington, DC: Author. Retrieved May 15, 2006, from http://www.census.gov/prod/www/abs/statab.html

Wainer, H., & Steinberg, L. S. (1992). Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: A bidirectional validity study. *Harvard Educational Review, 62*, 323-336.

Weisstein, N. (1970). Kinder, kuche, kirche as scientific law: Psychology constructs the female. In R. Morgan (Ed.), *Sisterhood is powerful: An anthology of writings from the women's liberation movement* (pp. 228-245). New York: Vintage Books.

Weisstein, N. (1997). Power, resistance and science. *New Politics, 6,* 145-167.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684.

Witelson, S. F., Beresh, H., & Kigar, D. L. (2005). Intelligence and brain size in 100 postmortem brains: Sex, lateralization and age factors. *Brain, 129*, 283-284.

Witt, S. D. (1997). Parental influence on children's socialization to gender roles. *Adolescence, 32*, 253-259.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358,* 749-750.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). The development of concern for others. *Developmental Psychology, 28,* 126-136.