Chapter 5

Meta-Analysis in the Psychology of Women

Janet Shibley Hyde
Shelly Grabe

The tradition of gender differences research has a long history in psychology, much of it predating the modern feminist movement and some of it clearly antifeminist in nature. In the late 1800s, for example, there was great interest in differences in the size of male and female brains and how they might account for the assumed lesser intelligence of women (Hyde, 1990; Shields, 1975). In the last several decades, the mass media and the general public have continued to be captivated by findings of gender differences. For example, John Gray's book *Men Are from Mars, Women Are from Venus* (1992), which argues for enormous psychological differences between women and men, has sold more than 30 million copies and has been translated into 40 languages (Gray, 2007). Deborah Tannen's book *You Just Don't Understand: Women and Men in Conversation* (1991) argues for the different-cultures hypothesis: that men's and women's patterns of speaking are so fundamentally different that they essentially belong to different linguistic communities or cultures. That book was on the *New York Times* best-seller list for nearly four years and has been translated into 24 languages (AnnOnline, 2007). Both works and dozens of others like them argue that males and females are, psychologically, vastly different. Yet as early as 1910, feminist researchers such as Helen Thompson Woolley wrote well-reasoned criticisms of the prevailing research.

A watershed book on psychological gender differences was Maccoby and Jacklin's *The Psychology of Sex Differences* (1974). Having reviewed

more than a thousand studies, they concluded that the following differences were fairly well established:

1. Girls have greater verbal ability than boys.
2. Boys outperform girls in spatial ability.
3. Boys perform better than girls on tests of mathematical ability.
4. Males are more aggressive.

They also challenged the long-standing traditional emphasis on gender differences and concluded that some beliefs in gender differences were unfounded, including such beliefs as:

1. Girls are more social than boys.
2. Girls are more suggestible (imitating and conforming).
3. Girls have lower self-esteem.
4. Girls are better at low-level cognitive tasks, boys at higher-level cognitive tasks.
5. Boys are more analytic.
6. Girls are more affected by heredity, boys by environment.
7. Girls have less achievement motivation.
8. Girls are more responsive to auditory stimuli, boys to visual stimuli.

That is, they noted many gender similarities. In the past several decades, feminist psychologists have become increasingly critical of the gender-differences tradition in psychological research. For example, some have argued that the emphasis on gender differences blinds us to gender similarities (Hyde, 1985; 2005).

In an important theoretical paper, Hare-Mustin and Marecek (1988) distinguished between "alpha bias" and "beta bias" in research and conceptualizations in the psychology of gender. *Alpha bias* refers to the exaggeration of gender differences. *Beta bias,* in contrast, refers to the minimizing of gender differences. From a feminist point of view, either can be problematic. If differences are exaggerated, for example, the research may serve as a basis for discrimination against women, who are "different." If real differences are minimized or ignored there are dangers, too; for example, if the large differences in men's and women's wages are ignored, divorce settlements might not provide adequate or equitable support for women and children (Weitzman, 1985).

Shortly after Maccoby and Jacklin's groundbreaking work in gender differences appeared, the statistical method of meta-analysis was developed (e.g., Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Rosenthal, 1991). This method revolutionized the study of psychological gender differences. Meta-analysis is a statistical technique that allows the researcher to synthesize results from numerous studies, and thus it is an especially appropriate tool to apply to questions of gender differences. Moreover, because it yields quantitative results—that is, it provides a measure of the magnitude of the gender difference—it can overcome problems of both alpha and beta bias. Modern techniques of meta-analysis also provide a highly nuanced view of gender differences, detecting, for example, those situations in which gender differences are more or less likely to be found. This chapter reviews existing meta-analyses of psychological gender differences. Following an introduction to the methods of meta-analysis, we review gender differences in cognitive performance, social behaviors, and motor behaviors.

META-ANALYTIC TECHNIQUES AND METHODOLOGICAL ISSUES

Traditional literature reviews—what might be called *narrative reviews*—are subject to several criticisms. They are nonquantitative, unsystematic, and subjective, and the task of reviewing 100 or more studies simply exceeds the information-processing capacities of the human reviewer (Hunter, Schmidt, & Jackson, 1982).

The review by Maccoby and Jacklin (1974) represented an advance because it made use of systematic vote counting. That is, Maccoby and Jacklin tabled all available studies of gender differences for a particular behavior, permitting the authors and the reader to count the number of studies finding a difference favoring females, the number finding a difference favoring males, and the number finding no difference.

The method of vote counting, unfortunately, also has flaws (Hedges & Olkin, 1985; Hunter et al., 1982). Statisticians have pointed out that vote counting can lead the reviewer to false conclusions (Hunter et al., 1982). For example, if there is a true gender difference in the population but the studies reviewed have poor statistical power (perhaps because of small sample sizes), the reviewer is likely to conclude that there is no effect because a majority of the studies may find no significant gender difference (for a detailed numerical example of this problem, see Hyde, 1986).

Statistical Methods in Meta-Analysis

Meta-analysis has been defined as the application of "quantitative methods to combining evidence from different studies" (Hedges & Olkin, 1985, p. 13). Essentially, then, it is a quantitative or statistical method for doing a literature review.

A meta-analysis proceeds in several steps. First, the researchers locate as many studies as they can on the particular question of interest. Computerized database searches are very useful in this phase. In the area of psychological gender differences, researchers can often obtain a very large sample of studies. For example, for a meta-analysis of gender differences in verbal ability, we were able to locate 165 studies reporting relevant data (Hyde & Linn, 1988).

Second, the researchers perform a statistical analysis of the statistics reported in each article. Crucial to meta-analysis is the concept of effect size, which measures the magnitude of an effect—in this case, the magnitude of gender difference. In gender meta-analyses, the measure of effect size typically is *d* (Cohen, 1988):

$$d = \frac{M_M - M_F}{S_w},$$

where M_M is the mean score for males, M_F is the mean score for females, and S_w is the average within-sex standard deviation. That is, *d* measures how far apart the male and female means are, in standardized units. Using this formula, negative values indicate higher average scores for females and positive values indicate higher average scores for males.

In meta-analysis, the effect sizes computed from all individual studies are then averaged to obtain an overall effect size reflecting the magnitude of gender differences across all studies. From a feminist point of view, one of the virtues of the *d* statistic is that it takes into account not only gender differences (the difference between male and female means), but also female variability and male variability (*s*, the standard deviation). That is, it recognizes that each sex is not homogenous.

If means and standard deviations for each sex are not available, *d* can be computed from other statistics, such as a *t*-test or *F*-test for gender differences. When the dependent variable is dichotomous (e.g., child fights or doesn't) and nonparametric statistics are used, they too can be converted to the effect size *d*. (For an excellent introduction to statistical methods in meta-analysis, see Lipsey & Wilson, 2001.)

In the third stage of the meta-analysis, the researchers average *d* values obtained from all studies. They can then reach conclusions such as: "Based on 165 studies that reported data on gender differences in verbal ability, the weighted mean effect size (*d*) was −0.11, indicating a slight female superiority in performance."

Meta-analytic methods make it possible to proceed one step further, to analyzing variations in values of *d*, that is, in the magnitude of the gender difference, according to various features of the studies (Lipsey & Wilson, 2001). This step is called *homogeneity analysis* because it analyzes the extent to which the values of *d* in the set are uniform or homogeneous. If there are large variations in the values of *d* across studies (and there invariably are), these variations reflect inconsistencies among the studies, and it is the task of the meta-analyst to account for the inconsistencies.

The meta-analysis then proceeds to a model-fitting stage. Either categorical or continuous models can be used. If a *categorical* model is used, the meta-analyst groups the studies into subsets or categories based on some logical, theoretically informed classification system. Statistically, the goal is to find a classification scheme that yields relatively homogenous values of d within each subset of studies. For example, in an analysis of gender differences in mathematics performance, one would compute an average value of d for studies that measured computation and another value of d for studies that measured mathematical problem solving. Thus investigators can determine whether the gender difference is large for some kinds of mathematics performance and close to zero for others, or even if the direction of the gender difference depends on the kinds of mathematics performance assessed—perhaps females perform better on some measures and males on others.

If a *continuous* model is used in the model-fitting stage, the meta-analyst uses a continuous variable, e.g., age, to account for variations among studies in the effect size, d. Eventually, a regression model is fitted in which the effect size is the criterion variable and some relevant continuous variable or variables are the predictors. For example, in studies of aggression, age may be a good predictor of the magnitude of the gender difference (Hyde, 1984).

METHODOLOGICAL ISSUES

A number of methodological issues in meta-analysis have been raised. Certainly chief among these is an issue of interpretation: When is an effect size large? Because of the way d is computed, it is a statistic much like z, and values can exceed 1. Thus it is impossible to say, in any absolute sense, that a value of 0.90, or any other value, is large. However, Cohen (1969) offered the following guidelines: a value of $d = 0.20$ is small, a value of 0.50 is moderate, and a value of 0.80 is large.

Rosenthal and Rubin (1982) introduced another scheme for deciding when an effect size is large. They used the Pearson correlation, r, rather than d, but the two can easily be translated using the approximation formula $d = 2r$ (or the exact formula

$$d = 2r\sqrt{(1-r^2)}.$$

To assess the magnitude of an effect size, they use the binomial effect size display (BESD). It displays the change in success rate (e.g., recovery from cancer due to treatment with a particular drug compared with an untreated control group) as a function of the effect size. For example, an r of 0.30 ($d = 0.60$) translates into an improvement in survival from 35 percent to 65 percent. Thus, according to Rosenthal and Rubin, effect sizes that appear only small to moderate may represent impressively large effects.

We would argue, however, that impressive effects in curing cancer do not necessarily transfer logically to the study of gender differences. In the latter case, the binomial effect size display can tell us something like the following: An effect size of $d = 0.40$ means that approximately 40 percent of one sex falls above the median (40 percent are above average) and 60 percent of the other sex falls above the median.

Another approach to interpreting the magnitude of an effect size is to compare it with effect sizes that have been obtained in other meta-analyses, either for related studies in the same field or for studies in other fields. One could compare the effect size for gender differences in mathematics performance with the effect size for gender differences in spatial ability, for example. Or one might compare the effect size for social class or ethnic differences in math performance. Table 5.1 provides effect sizes for gender differences documented in numerous meta-analyses.

Another major methodological issue in meta-analysis concerns the sampling of studies and the potential for sampling bias. Ideally, the sampling procedure should be well defined, systematic, and exhaustive. A poor sampling procedure will produce misleading, if not useless, results. Even with good sampling procedures, however, problems can arise because studies that found significant effects are more likely to be published than those that did not. This biases the published results in the direction of larger effect sizes. In addition, investigators may not publish data that show large and significant effects that run counter to the zeitgeist, a tendency that would serve to maintain a status quo in the literature. One way of guarding against sample bias is to seek out unpublished studies. Doctoral dissertations and major national data sets such as the National Longitudinal Survey of Youth (NLSY) or National Assessment of Educational Progress (NAEP) are perhaps the best sources of unpublished data that may show nonsignificant effects or failures to replicate.

A final issue concerns the validity of meta-analytic research on gender differences. As Eagly (1986) points out, both the construct and external validities of the aggregated results of a meta-analysis are probably greater than those of most individual studies. However, threats to that greater validity do exist and cannot be ignored. To the extent that studies in the sample rely on similar measurement instruments or have other features in common, validity may be compromised. Examples are stimulus materials that inadvertently favor one gender over the other, samples that are unrepresentative of the population, and a preponderance of laboratory, as opposed to field, studies. Eagly (1986) recommended using meta-analytic techniques to assess the effects of those study characteristics.

Table 5.1.
Major meta-analyses of research on psychological gender differences

Study	Variable	Age	Number of Reports	*d*
Cognitive Variables				
Hyde, Fennema, & Lamon (1990)	Mathematics computation	All ages	45	−0.14
	Mathematics concepts	All ages	41	−0.03
	Mathematics problem solving	All ages	48	+0.08
Hedges & Nowell (1995)	Reading comprehension	Adolescents	5*	−0.09
	Vocabulary	Adolescents	4*	+0.06
	Mathematics	Adolescents	6*	+0.16
	Perceptual speed	Adolescents	4*	−0.28
	Science	Adolescents	4*	+0.32
	Spatial ability	Adolescents	2*	+0.19
Hyde, Fennema, Ryan, et al. (1990)	Mathematics self-confidence	All ages	56	+0.16
	Mathematics anxiety	All ages	53	−0.15
Feingold (1988)	DAT Spelling	Adolescents	5*	−0.45
	DAT Language	Adolescents	5*	−0.40
	DAT Verbal reasoning	Adolescents	5*	−0.02
	DAT Abstract reasoning	Adolescents	5*	−0.04
	DAT Numerical ability	Adolescents	5*	−0.10
	DAT Perceptual speed	Adolescents	5*	−0.34

Study	Variable	Age	Number of Reports	*d*
	DAT Mechanical reasoning	Adolescents	5*	+0.76
	DAT Space relations	Adolescents	5*	+0.15
Hyde & Linn (1988)	Vocabulary	All ages	40	−0.02
	Reading comprehension	All ages	18	−0.03
	Speech production	All ages	12	−0.33
Linn & Petersen (1985)	Spatial perception	All ages	62	+0.44
	Mental rotation	All ages	29	+0.73
	Spatial visualization	All ages	81	+0.13
Voyer et al. (1995)	Spatial perception	All ages	92	+0.44
	Mental rotation	All ages	78	+0.56
	Spatial visualization	All ages	116	+0.19
Lynn & Irwing (2004)	Progressive matrices	6–14 years	15	+0.02
		15–19 years	23	+0.16
		Adults	10	+0.30
Whitley et al. (1986)	Attribution of success to ability	All ages	29	+0.13
	Attribution of success to effort	All ages	29	−0.04
	Attribution of success to task	All ages	29	−0.01
	Attribution of success to luck	All ages	29	−0.07
	Attribution of failure to ability	All ages	29	+0.16
	Attribution of failure to effort	All ages	29	+0.15
	Attribution of failure to task	All ages	29	−0.08
	Attribution of failure to luck	All ages	29	−0.15
Rosenthal & Rubin (1982)	Verbal ability	11 years and older	12	−0.30

(Continued)

Table 5.1.
Major meta-analyses of research on psychological gender differences (*Continued*)

Study	Variable	Age	Number of Reports	d
	Quantitative ability	11 years and older	7	+0.35
	Visual-spatial ability	11 years and older	7	+0.50
	Field articulation	11 years and older	14	+0.51
Communication				
Anderson & Leaper (1998)	Interruptions in conversation	Adults	53	+0.15
	Intrusive interruptions	Adults	17	+0.33
Leaper & Smith (2004)	Talkativeness	Children	73	−0.11
	Affiliative speech	Children	46	−0.26
	Assertive speech	Children	75	+0.11
Dindia & Allen (1992)	Self-disclosure, all studies	Not reported	205	−0.18
	Self-disclosure to stranger	Not reported	99	−0.07
	Self-disclosure to friend	Not reported	50	−0.28
LaFrance et al. (2003)	Smiling	Adolescents and adults	418	−0.40
	Smiling: Aware of being observed	Adolescents and adults	295	−0.46
	Smiling: Not aware of being observed	Adolescents and adults	31	−0.19
McClure (2000)	Facial expression processing	Infants	29	−0.18 to −0.92
	Facial expression processing	Children and adolescents	89	−0.13 to −0.18
Hall & Halberstadt (1986)	Social smiling	Children	5	+0.04
	Social smiling	Adults	15	−0.42
	Social gazing	Children	11	−0.48
	Social gazing	Adults	30	−0.69
Stier & Hall (1984)	Initiate touch	All ages	6	−0.09
	Receive touch	All ages	5	+0.02
Social Variables				
Hyde (1984, 1986)	Aggression (all types)	All ages	69	+0.50
	Physical aggression	All ages	26	+0.60
	Verbal aggression	All ages	6	+0.43
Eagly & Steffen (1986)	Aggression	Adults	50	+0.29
	Physical aggression	Adults	30	+0.40
	Psychological aggression	Adults	20	+0.18
Knight et al. (2002)	Physical aggression	All ages	41	+0.59
	Verbal aggression	All ages	22	+0.28
	Aggression in low emotional arousal context	All ages	40	+0.30
	Aggression in emotional arousal context	All ages	83	+0.56
Bettencourt & Miller (1996)	Aggression under provocation	Adults	57	+0.17
	Aggression under neutral conditions	Adults	50	+0.33
Archer (2004)	Aggression in real-world settings	All ages	75	+0.30 to +0.63

(*Continued*)

Table 5.1.
Major meta-analyses of research on psychological gender differences (*Continued*)

Study	Variable	Age	Number of Reports	d
	Physical aggression	All ages	111	+0.33 to +0.84
	Verbal aggression	All ages	68	+0.09 to +0.55
	Indirect aggression	All ages	40	−0.74 to +0.05
Stuhlmacher & Walters (1999)	Negotiation outcomes	Adults	53	+0.09
Walters et al. (1998)	Negotiator competitiveness	Adults	79	+0.07
Eagly & Crowley (1986)	Helping behavior	Adults	99	+0.13
	Helping: Surveillance context	Adults	16	+0.74
	Helping: No surveillance	Adults	41	−0.02
Oliver & Hyde (1993)	Sexuality: Masturbation	All ages	26	+0.96
	Sexuality: Attitudes about casual sex	All ages	10	+0.81
	Sexual satisfaction	All ages	15	−0.06
	Attitudes about extramarital sex	All ages	17	+0.29
Murnen & Stockton (1997)	Arousal to sexual stimuli	Adults	62	+0.31
Eagly & Johnson (1990)	Leadership: Interpersonal style	Adults	153	−0.04 to −0.07
	Leadership: Task style	Adults	154	0.00 to −0.09
	Leadership: Democratic vs. autocratic	Adults	28	+0.22 to +0.34
Eagly et al. (1992)	Leadership: Evaluation	Adults	114	+0.05
Eagly et al. (1995)	Leadership effectiveness	Adults	76	−0.02
Eagly et al. (2003)	Leadership: Transformational	Adults	44	−0.10
	Leadership: Transactional	Adults	51	−0.13 to +0.27
	Leadership: Laissez-faire	Adults	16	+0.16
Feingold (1994)	Neuroticism: Anxiety	Adolescents and adults	13*	−0.32
	Neuroticism: Impulsiveness	Adolescents and adults	6*	−0.01
	Extraversion: Gregariousness	Adolescents and adults	10*	−0.07
	Extraversion: Assertiveness	Adolescents and adults	10*	+0.51
	Extraversion: Activity	Adolescents and adults	5	+0.08
	Openness	Adolescents and adults	4*	+0.19
	Agreeableness: Trust	Adolescents and adults	4*	−0.35
	Agreeableness: Tendermindedness	Adolescents and adults	10*	−0.91
	Conscientiousness	Adolescents and adults	4	−0.18
Wood (1987)	Individual performance	Adults		+0.38
Psychological Well-Being				
Kling et al. (1998) Analysis I	Self-esteem	All ages	216	+0.21
Kling et al. (1998) Analysis II	Self-esteem	Adolescents	15*	+0.04 to +0.16
Major (1999)	Self-esteem	All ages	226	+0.14
Feingold & Mazzella (1998)	Body esteem	All ages	NA	+0.58

(*Continued*)

Table 5.1.
Major meta-analyses of research on psychological gender differences (*Continued*)

Study	Variable	Age	Number of Reports	d
Twenge & Nolen-Hoeksema (2002)	Depression symptoms	8–16 years	310	+0.02
Wood et al. (1989)	Life satisfaction	Adults	17	–0.03
	Happiness	Adults	22	–0.07
Pinquart & Sörensen (2001)	Life satisfaction	Elderly	176	+0.08
	Self-esteem	Elderly	59	+0.08
	Happiness	Elderly	56	–0.06
Tamres et al. (2002)	Coping: Problem-focused	All ages	22	–0.13
	Coping: Rumination	All ages	10	–0.19
Motor Behaviors				
Thomas & French (1985)	Balance	3–20 years	67	+0.09
	Grip strength	3–20 years	37	+0.66
	Throw velocity	3–20 years	12	+2.18
	Throw distance	3–20 years	47	+1.98
	Vertical jump	3–20 years	20	+0.18
	Sprinting	3–20 years	66	+0.63
	Flexibility	5–10 years	13	–0.29
Eaton & Enns (1986)	Activity level	All ages	127	+0.49
Miscellaneous				
Thoma (1986)	Moral reasoning: Stage	Adolescents and adults	56	–0.21

Study	Variable	Age	Number of Reports	d
Jaffee & Hyde (2000)	Moral reasoning: Justice orientation	All ages	95	+0.19
	Moral reasoning: Care orientation	All ages	160	–0.28
Silverman (2003)	Delay of gratification	All ages	38	–0.12
Whitley et al. (1999)	Cheating behavior	All ages	36	+0.17
	Cheating attitudes	All ages	14	+0.35
Whitley (1997)	Computer use: Current	All ages	18	+0.33
	Computer self-efficacy	All ages	29	+0.41
Konrad et al. (2000)	Job attribute preferences:			
	Earnings	Adults	207	+0.12
	Security	Adults	182	–0.02
	Challenge	Adults	63	+0.05
	Physical work environment	Adults	96	–0.13
	Power	Adults	68	+0.04

Note: Positive values of *d* represent higher scores for males. NA means not available; article did not provide this information clearly. DAT means Differential Aptitude Test. *Based on major, large national samples.
Source: J. S. Hyde, 2005. The gender similarities hypothesis. *American Psychologist, 60*, 581–592. Reprinted with permission.

META-ANALYSIS AND GENDER DIFFERENCES IN COGNITIVE PERFORMANCE

Verbal Abilities

One supposed gender difference is in verbal ability. Hyde and Linn (1988) meta-analyzed 165 reports of gender differences in verbal ability, 120 of which reported data adequate for effect size computations. Three-fourths of the *d* values were negative, and the mean value was −0.11, indicating a slight female superiority. Homogeneity analyses revealed that *d* varied with type of verbal ability (mean *d* was −0.02 for vocabulary, 0.16 for analogies, −0.03 for reading comprehension, −0.33 for speech production, −0.09 for essay writing, −0.22 for anagrams, and −0.20 for general verbal ability). In light of these findings, Hyde and Linn concluded that the magnitude of the gender difference in verbal ability is "effectively zero" (p. 64).

Spatial, Science, and Quantitative Abilities

Linn and Petersen (1985) focused on spatial ability in their meta-analysis. They culled 172 independent effect sizes from their sample and assigned each of them to one of three categories of spatial ability. For spatial perception (defined as the ability to determine spatial relationships with respect to one's own orientation), they found a mean effect size of 0.44, indicating better male performance. For mental rotation, the value was 0.73. For spatial visualization (defined as the ability to perform complex, multistep spatial manipulations), it was 0.13. These heterogeneous results render as inappropriate all global statements about gender differences in spatial ability.

Linn and Petersen analyzed their data for age trends in the magnitude of the effect sizes. They wanted to assess the evidence for the argument that gender differences in spatial ability are biologically based because they emerge in adolescence. Their results did not support this hypothesis. For example, the mean *d* for studies of spatial perception in persons under the age of 13 was the same as the mean *d* for the studies of spatial perception in persons between the ages of 13 and 18 (in each case, mean $d = 0.37$). Of course, these results do not resolve the issue of the origin of gender differences in spatial ability because not all biological explanations posit a pubertal onset.

In a more recent meta-analysis of gender differences in spatial abilities Voyer, Voyer, and Bryden (1995) analyzed 286 effect sizes and reported an overall mean weighted *d* of 0.37, demonstrating gender differences in overall spatial abilities that favor males. Homogeneity analyses using the same categories employed by Linn and Peterson (1985) indicated that mean effect sizes for spatial perception ($d = 0.44$), mental rotation ($d = 0.56$), and spatial visualization ($d = 0.19$) were

comparable or smaller. Voyer et al. further demonstrated that the reported gender differences were moderated by age. Specifically, effect size magnitude increased with age for each outcome: spatial perception ($d = 0.33$, under 13 years; $d = 0.43$, 13–18 years; and $d = 0.48$, over 18 years), mental rotation ($d = 0.33$, under 13 years; $d = 0.45$, 13–18 years; and $d = 0.66$, over 18 years), and spatial visualization ($d = 0.02$, under 13 years; $d = 0.18$, 13–18 years; and $d = 0.23$, over 18 years).

Hyde, Fennema, and Lamon (1990) meta-analyzed 100 studies of mathematics performance, assessing the evidence for the effects of gender, task, and age. Across studies of samples of the general population, they obtained an average value of −0.05, indicating a negligible female advantage. An analysis of age trends revealed that females outperform males in computation in both elementary ($d = -0.20$) and middle school ($d = -0.22$) and that males outperform females in problem solving in high school ($d = 0.29$) and college ($d = 0.32$). Hyde and colleagues also found an effect for sample selectivity, in that studies of highly selective or precocious populations produced the largest gender differences. Finally, they provided evidence that cognitive gender differences are getting smaller: the mean effect size for studies published before 1974 was 0.31, whereas the mean *d* value for later studies was 0.14. Hyde et al. argued that Maccoby and Jacklin's (1974) conclusion that "boys excel in mathematical ability" (p. 352) is oversimplified and is by now outdated. This meta-analysis used mathematics performance on standardized tests as the measure. If one looks instead at math grades in school, girls perform better than boys at all grade levels (Kimball, 1989).

Mathematics Attitudes and Affect

Hyde and her colleagues (Hyde, Fennema, Ryan, Frost, & Hopp, 1990) examined 70 reports of gender differences in mathematics attitudes and affect. The dependent variables included mathematics anxiety, mathematics self-concept, parental attitudes toward the child's participation in mathematics, and mathematics success and failure attributions. The effects on *d* of the age of the children, the year of publication, and the selectivity of the sample were evaluated.

Hyde and colleagues found mostly small effect sizes (more than half were one-tenth of a standard deviation or less) for all age groups combined. The one exception to this pattern was the stereotyping of math as a male domain. It yielded a large effect size (mean $d = -0.90$), indicating that males stereotype mathematics as a masculine activity more than females do. Homogeneity analyses revealed that this gender difference in stereotyping—as well as gender differences (that favored boys) in parents' and teachers' attitudes toward the subject's participation in mathematics—peaks in the high school years. The size of the gender difference in mathematics anxiety was associated with the selectivity of the sample: it was lowest in the highly selected, precocious samples (mean $d = 0.09$) and highest in the remedial and math anxiety classes (mean $d = 0.30$).

Regression analyses showed that male students reported more positive parental and teacher attitudes in the 1970s but that female students reported more positive attitudes in the 1980s and that the gender difference in stereotyping of mathematics as a male domain has decreased somewhat over time. The authors urged caution in interpreting the former result, however, because one cannot tell from the data whether the attitudes of significant adults had become more positive toward girls or more negative toward boys.

Overall, Hyde et al. concluded that gender differences in mathematics attitudes and affect are small—too small to account for women's underrepresentation in mathematics-related occupations (thus urging us to look elsewhere for an explanation), but not so small that they can be ignored (the cumulative effect of many small disadvantages for females may still be a powerful one).

META-ANALYSES OF GENDER DIFFERENCES IN SOCIAL BEHAVIOR

Aggression

Hyde (1984, 1986) meta-analyzed a set of 143 studies of gender differences in aggression. A mean *d* value of 0.50 was obtained for 69 general samples. Hyde found a significant age trend in the data, indicating that the gender difference in aggression varied inversely with the average age of subjects in the study. That is, gender differences in aggression were larger among preschoolers (median $d = 0.58$) and smaller among college students (median $d = 0.27$).

Using Hedges's (1982a, 1982b) homogeneity statistics, Hyde found that type of research design (i.e., experimental versus naturalistic), method of measurement (e.g., direct observation, self-report, parent or teacher report), and type of aggression sampled (e.g., physical, verbal) all produced significant between-category differences. The naturalistic/correlational studies yielded significantly larger gender differences in aggression than did the experimental studies (mean $d = 0.56$ versus mean $d = 0.29$). However, she did not find significant differences between studies of physical aggression and studies of verbal aggression.

At about the same time that Hyde's work appeared, Eagly and Steffen (1986) published a meta-analysis of gender differences in aggression that had been reported in the experimental social psychological literature. They restricted their sample to studies of persons 14 years of age and older (most were college-age samples) and to studies in which the dependent variable was a behavioral measure of aggression toward another person. These restrictions resulted in a fairly homogeneous group of laboratory and field studies in which relatively brief encounters with strangers were assessed.

The sample of 63 studies yielded 50 independent effect sizes for analysis. Across all 50 values, the mean weighted effect size was 0.29, indicating greater male aggressiveness. However, heterogeneity analyses revealed that the mean *d* was greater for the laboratory (0.35) than for the field studies (0.21) and greater for studies of physical (0.40) than psychological (0.18) aggression. They also found that the gender difference was larger for semiprivate than for public experimental settings (0.38 versus 0.17). Also of note is the fact that every mean effect size calculated was positive, indicating great consistency in the direction of the gender difference (even though there is clearly great inconsistency in its size).

As part of their effort to fit continuous models to their effect size data, Eagly and Steffen had 200 undergraduates rate brief descriptions of the aggressive behaviors described in the studies in their sample for (1) harmfulness to the target, (2) anxiety/guilt for the respondent, and (3) dangerousness for the respondent. The participants were also asked how likely they thought it to be that (1) they, (2) the average woman, and (3) the average man would enact the aggressive behavior. The group's responses to these six questions, scored for gender differences, were included in the set of predictor variables used in the regression-type analysis. The gender difference in the undergraduate respondents' assessment of how much anxiety/guilt and danger they would feel had they perpetrated such an act of aggression predicted the magnitude of gender differences. That is, to the extent that the women respondents reported that they would feel more anxiety/guilt and danger in that situation than the men reported they would feel, *d* was large. The results of this set of analyses were, by and large, interpretable within the framework of Eagly's (1986) social role theory.

Although the Hyde et al. and Eagly and Steffen meta-analyses have shown gender differences that are moderate in magnitude, the gender difference in physical aggression is more reliable and larger than the gender difference in verbal aggression. Based on a later meta-analysis of gender differences in aggression, Archer (2004) reported that indirect or relational aggression showed an effect size for gender differences of −0.45 when measured by direct observation (just 3 studies), but was only −0.19 for peer ratings (14 studies), −0.02 for self-reports (40 studies), and −0.13 for teacher reports (8 studies). Therefore, evidence is ambiguous regarding the magnitude of the gender difference in relational aggression.

Helping Behaviors

A meta-analysis of research on gender differences in helping behavior was performed by Eagly and Crowley (1986) that is as deeply rooted in social theory as is Eagly and Steffen's (1986) work on gender differences in aggression. Eagly and Crowley were able to cull 99 effect sizes from the 172 studies they found. The mean weighted effect size was 0.34, indicating greater helping behavior among men. This result seems, at first, counterintuitive, because helping is central to the female role. However, it is exactly what social role theory predicts. The key to understanding this result is an appreciation of the dynamics of the typical social psychological study of helping behavior (which was the only type of study Eagly and Crowley included in their sample). The studies examined relatively brief encounters with strangers, encounters that call for "chivalrous acts and nonroutine acts of rescuing" (p. 300). As Eagly and Crowley argued convincingly, these are exactly the types of helping behaviors that the male role fosters. The female role, in contrast, fosters caretaking and helping behaviors primarily in the context of ongoing close relationships, which are not assessed in psychologists' typical research.

The results indicated that the gender difference in helping behavior was larger (in the male direction) in off-campus settings than in the laboratory, when there were other people around to witness the act than when there were not, when other helpers were available than when there were not, and when the appeal for help was a presentation of a need rather than a direct request. The results indicated that larger effect sizes (again, in the male direction) were associated with gender differences in the undergraduate raters' reports of how competent, comfortable, and endangered they would feel performing the helping behavior. In other words, to the extent that the male undergraduate raters said they would be more likely to perform the helping behavior and feel more competent, more comfortable, and less endangered doing it than did the female, the behavior was associated with a larger gender difference.

As mentioned earlier, Eagly and Crowley also analyzed the target—or requester—gender effects. Across 36 values, the mean weighted effect size was −0.46, indicating that women received more help than men did. The correlation between the effect size for the target's gender and the effect size for the participant's gender was negative and significant, $r = -0.40$. Thus, not surprisingly, the study characteristics that related significantly to subject gender effect size (i.e., setting surveillance, availability of helpers, type of request) were also related to target gender effect size, though in the opposite direction. Further analysis of these data revealed that men were more likely to help women than men, but received help from men and women about the same; whereas women were equally likely to help men and women, but more often received help from men than from women.

Small Group Behavior

Wood (1987) focused her meta-analysis on gender differences in group productivity. She restricted her review to laboratory studies in which an objective measure of performance on the assigned task was used. The 52 studies she found were coded for whether group members worked on the task individually or together, how it was scored (for creativity, number of solutions, time to completion, number of errors, and so on), and whether it required task-oriented or social activity for better performance. Wood found that men outperformed women when working individually in same-sex groups (mean effect size of 0.38 across 19 values). She found no evidence of a gender difference in individual performance while working in mixed-sex groups (5 studies) and only a significant tendency for mixed groups to outperform single-sex groups of either gender (8 studies).

Wood's categorical model-fitting analyses (done only on the same-sex data) yielded just two significant effects. First, for the dependent measure of number of solutions, there was better male performance when group members worked alone (mean $d = 0.78$), but not when they worked together (mean $d = -0.05$). That is, men generated more solutions than did women when they worked alone in same-sex groups, but the two sexes generated equal numbers of solutions when they worked in groups together with other members of their own gender. Second, on tasks that require task-oriented behavior for good performance, men outperformed women whether they were working individually (mean $d = 0.25$) or together (mean $d = 0.34$), whereas on tasks that require social behavior for good performance, women performed slightly better (mean $d = -0.11$). Study variables that accounted for small, but significant, portions of the variance in gender effect size were male authorship and more recent year of publication: a greater percentage of male authors and more recent year of publication were associated with larger effects. Wood called for greater appreciation, in the workplace, of the specific facilitative effects of women's interaction style on group productivity.

Leadership Behavior

Eagly and colleagues, across several meta-analyses, have thoroughly reviewed the leadership literature. Eagly and Johnson (1990) evaluated gender difference in autocratic versus democratic (also known as directive versus participative) leadership style, as well as in task versus interpersonal orientation. The 144 studies in their analysis included laboratory experiments, assessment studies, and field studies in organizational settings. Because of their belief that, in real-life settings, male and female leaders are selected according to the same criteria, Eagly and Johnson predicted that they would find smaller gender differences in the field studies than in the other two types of reports. Their prediction was supported by their results.

Across all 329 effect sizes, Eagly and Johnson obtained a mean values of 0.03, indicating virtually no gender difference. They found similarly near-zero mean effect sizes across gender comparisons on interpersonal style measures, task style measures, and bipolar measures that assessed the two styles simultaneously. However, they found a more substantial gender difference for democratic versus autocratic style (mean $d = -0.22$), a finding that suggests women are more democratic than men in their leadership style.

When they looked at the three types of studies (organizational, assessment, laboratory) in their sample separately, Eagly and Johnson found strong support for their major prediction regarding field studies, as well as consistent evidence for a gender difference in democratic versus autocratic style. More specifically, across the effect sizes computed from the 269 organizational studies, they obtained mean values of 0.01, −0.02, 0.03, and −0.21 for interpersonal style, task style, interpersonal style versus task style, and democratic versus autocratic style, respectively. The analogous values for the 43 assessment studies were −0.25, 0.08, 0.04, and −0.29; and the values for the 17 laboratory studies were −0.37, 0.19, −0.12, and −0.20. Thus, with the exception of democratic versus autocratic style, larger gender differences were obtained in studies of persons who do not actually occupy leadership positions and who are evaluated in artificial and contrived settings. In these studies, men behave in a more task-oriented fashion and women, in a more interpersonally oriented one.

The tendency for women to lead in a more democratic way and men to do so in a more autocratic way, in contrast, is found across all types of studies. Indeed, the authors found that 92 percent of the gender comparisons on this dimension were in the stereotypic direction. Eagly and Johnson suggested that female and male leaders bring to their leadership positions a wealth of gender-based experience. Consequently, though they may be selected according to the same criteria, they are not equivalent persons. Eagly and Johnson also suggested that female leaders may attempt to placate their coworkers by asking for their input, in order to cope with continued institutional hostility toward women leaders. Lastly, although Eagly and Johsnon did not argue for the greater effectiveness of a participative leadership style, they did note the current trend away from rigid, hierarchical management practices, a trend presumably guided by that belief.

In a more recent meta-analysis examining contemporary leadership styles, Eagly, Johannesen-Schmidt, and van Engen (2003) reviewed research that compared women and men on transformational, transactional, and laissez-faire leadership styles. The meta-analysis of 45 studies found that, on average, female leaders were slightly more transformational than male leaders in their leadership ($d = -0.10$). Predicted gender-related differences were also found when the transformational and transactional scales of the Multifactor Leadership Questionnaire were broken down into their respective subscales. For

example, it was found that women scored higher than men on the transformational subscale of Individualized Consideration ($d = -0.23$). Men scored higher than women on one of the transactional subscales, Management by Exception-Passive ($d = 0.27$), whereas women scored slightly higher on the Contingent Reward subscale ($d = -0.13$). Men also scored higher on laissez-faire leadership ($d = 0.16$). The overall comparisons on transformational leadership, as well as its subscales, show significantly higher scores among women than men, whereas men obtained significantly higher scores on management by exception and laissez-faire styles.

Interestingly, the authors found that the reported gender differences in leadership style were moderated by setting and publication year. In particular, the authors found the smallest differences in business settings ($d = -0.07$), as opposed to governmental ($d = -0.11$) or educational ($d = -0.21$) settings. Furthermore, when publication year was taken into account, findings revealed that the gender difference reported in transformational style has gone more strongly in the female direction in recent years. Over time, perhaps women have perceived less pressure to conform to a traditionally masculine style of leadership and have experienced more freedom to lead in a manner that they are comfortable with. However, the small effects suggest that although there are differences in leadership styles between women and men, they are not large.

Another rich area of research that examines gender-related differences in leadership is the investigation of the relative effectiveness of men and women who occupy leadership roles in groups or organizations. Eagly, Karau, and Makhijani (1995) reviewed 76 studies that compared women and men managers, supervisors, officers, department heads, and coaches. Effectiveness was measured by subjective ratings anchored by *poor* leader and *outstanding* leader. When all studies in the literature were aggregated, female and male leaders did not differ in effectiveness ($d = -0.02$). However, although the overall finding indicated men and women were equivalent in effectiveness, that generalization was not appropriate in all organizational contexts. In particular, follow-up analyses indicated that findings from studies that investigated military organizations differed from the rest. When military organizations were excluded from analyses, the weighted mean effect size indicated that female leaders were rated as slightly more effective than male leaders ($d = -0.12$).

The magnitude of the overall effect size also was moderated by the traditional masculinity of the role and the sex of the subordinates. Comparisons of leader effectiveness favored men more and women less to the extent that the leadership role was male-dominated and that the subordinates were male. Recall that if military studies are included there was no overall gender difference. The remaining small and insignificant difference is important because it suggests that despite barriers and possible challenges in leadership, the women who serve as leaders are in general succeeding as well as their male counterparts. Similarly, despite the meta-analytic findings reviewed earlier that suggest that female leaders appear to behave somewhat differently than male leaders, these findings suggest that they appear to be equally effective. Furthermore, even though the data suggest that men may excel in some areas and women may excel in others, there appears no empirical reason to believe that either gender possesses an overall advantage in effectiveness.

Because gender stereotypes may cause behavior to be interpreted differently for female leaders, the issue of leadership evaluation is also important. Eagly, Makhijani, and Klonsky's (1992) synthesis of 147 experiments that examined evaluations of female and male leaders whose behavior had been made equivalent by the researchers found that evaluations were less favorable for female than for male leaders, but the effect size ($d = 0.05$) was so small that a conclusion of no effect seems reasonable. However, the bias for female leaders to be devalued was larger in specific contexts. Female leaders were devalued relative to their male counterparts when they adopted equivalent leadership styles that were stereotypically masculine (i.e., an autocratic and directive style) as well as when their evaluators were men. In contrast, female and male leaders were evaluated favorably when they adopted equivalent leadership styles that were traditionally feminine (i.e., democratic or interpersonally oriented). The finding that devaluation of women in leadership roles was stronger when leaders occupied male-dominated roles and when their evaluators were men suggests that women's occupancy of highly male-dominated leadership roles produces a violation of people's expectancies about women. Male evaluators may experience female leaders as a more threatening intrusion because leadership is traditionally a male domain.

The authors also found that the tendency to favor men over women was larger when the dependent variable was the leader's competence or rater's satisfaction with the leader rather than the perception of leadership style. Thus, the measures that were more purely *evaluative* (i.e., competence or satisfaction) yielded stronger evidence of the devaluation of women's leadership. When specific leadership style was the moderator, two of three styles examined (interpersonal orientation and potency) did not produce gender differences. However, women were perceived as more task-oriented than men. This perception, contrary to what would be expected, may reflect a tendency to view women's behavior as more extreme when it conflicts with the female stereotype. The autocratic leadership style produced significantly more favorable evaluations of male than female leaders ($d = 0.30$), but only trivial differences were found for roles occupied mainly by men ($d = 0.09$) than for those occupied equally by men and women ($d = -0.06$). There was a greater tendency to favor male leaders in male-dominated leadership positions of business and manufacturing than in organizational contexts not involving business or manufacturing. These results highlight that men's styles may be less consequential in that their leadership is not questioned and they therefore enjoy greater latitude to carry out leadership in a variety of styles.

Nonverbal Communication

Stier and Hall (1984) reviewed 43 observational studies of gender differences in touch and obtained a complex and somewhat ambiguous pattern of results. Looking first at the direction of the findings, they found that 63 percent of the studies reported more female-to-male than male-to-female touching, 71 percent reported more female-to-female than male-to-male touching, 64 percent reported more touch initiated by females, and 61 percent reported more touch received by females. However, the average effect sizes associated with each of these four variables were all near zero (0.02, 0.00, −0.09, and 0.02, respectively). Stier and Hall also reported that the majority of studies found that females react more favorably to touch than do males, although they did not include an average effect size. Their failure to find clear-cut evidence for an asymmetry in touching behavior in opposite-gender dyads forced Stier and Hall to conclude that Henley's (1977) power hypothesis did not have a strong empirical base. They did, however, suggest a modification. Drawing on Goldstein and Jefford's (1981) finding that lower-status legislators touched higher-status legislators more often than the other way around, Stier and Hall speculated that touching may be more consistent with lower, rather than higher, status and reflect the individual's "strong desire either to redress the status imbalance or to establish a bond of solidarity" (p. 456).

In her review of the literature on gender differences in nonverbal communicative behaviors, Hall (1984) devoted a chapter to each of the following topics, quantifying the evidence wherever possible: interpersonal sensitivity and judgment accuracy, expression accuracy, facial behavior, gaze, interpersonal distance and orientation, touch, body movement and position, and voice. In her concluding chapter, she provided a table (table 11.1, p. 142) in which the average point-biserial correlations between gender and performance for 21 nonverbal behaviors are displayed. (To obtain a rough comparability of statistics, $d = 2r$.) Each average effect size is based on at least five independent studies, and, where they exist, separate results are reported for infants, children, and adolescents. The data indicate that women are better at decoding nonverbal communication ($r = -0.21$), recognizing faces ($r = -0.17$), and expressing emotions nonverbally ($r = -0.25$); that they have more expressive faces ($r = -0.45$), smile ($r = -0.30$) and gaze ($r = -0.32$) more, receive more gaze ($r = -0.32$), approach ($r = -0.27$) and are approached by others more closely ($r = -0.43$), and make fewer speech errors ($r = 0.33$) and filled pauses ($r = 0.51$); and that their body movements are less restless ($r = 0.34$), less expansive ($r = 0.46$), more involved ($r = -0.16$), more expressive ($r = -0.28$), and more self-conscious ($r = -0.22$). Surely it was this set of results that led Hall and Halberstadt (1986) to comment, two years later, "In sum, based on a literature of hundreds of studies, it appears that women occupy a more nonverbally conscious, positive, and interpersonally engaged world than men do" (p. 137).

In a recent meta-analysis of research on gender differences in smiling, LaFrance, Hecht, and Paluck (2003) analyzed 418 samples and found a moderate difference ($d = -0.41$), with girls and women smiling more. However, the authors reported that the observed gender difference was highly dependent on context: if participants had a clear awareness that they were being observed, the gender difference was larger ($d = -0.46$) than if they were not aware of being observed ($d = -0.19$). The magnitude of the gender difference also depended on age and culture. Gender differences were largest among adolescents ($d = -0.56$, 13–17 years), smaller among young adults ($d = -0.45$, 18–23 years), small during adulthood ($d = -0.30$, 24–64 years), and near zero after age 65 ($d = -0.11$). Interestingly, gender differences were largest among Caucasian samples ($d = -0.43$) and smaller and comparable among African American, Native American, Indian, Asian, Australian Aboriginal, or "mixed" samples ($d = -0.25, -0.27, -0.37, -0.30, -0.22$, and -0.34, respectively).

META-ANALYSIS AND GENDER DIFFERENCES IN PSYCHOLOGICAL WELL-BEING

Taylor and Hall (1982) conducted a meta-analytic review of 107 reports of the effects of masculinity and femininity on self-esteem, adjustment, ego development, and other measures of mental health. They carried out their analysis in the context of a theoretical reconceptualization of androgyny within the framework of a two-way analysis of variance. According to this approach, Bem's (1974) model of androgyny predicts a significant interaction, whereas Spence, Helmreich, and Stapp's (1974) model predicts significant main effects for both masculinity and femininity.

Across all 107 reports, Taylor and Hall found that the strength of the association between masculinity and mental health was stronger than that between femininity and mental health, both for each gender and for each type of dependent measure. For example, the average correlation between masculinity and adjustment was 0.53 for men and 0.31 for

women, whereas the average correlation between femininity and adjustment was 0.05 for men and 0.04 for women. In addition, Taylor and Hall found that, of the results that addressed the issue, about half favored psychologically balanced individuals and half favored sex-typed individuals. Taylor and Hall concluded that the traditional notion that feminine women and masculine men embody psychological health clearly must be rejected and that the balance model of androgyny has minimal and inconsistent empirical support. Rather, they argued, for each gender, "it is primarily masculinity that pays off" (p. 362).

Wood, Rhodes, and Whelan (1989) conducted a meta-analytic review of 93 studies of gender differences in life satisfaction and well-being. They were particularly interested in the effects associated with marriage, which they predicted would be especially salutary for women. Because studies of life satisfaction tend to be done disproportionately on elderly and disabled persons, Wood et al. ran validation analyses on a subset of 18 studies with samples that were representative of the U.S. population. Across the 85 effect sizes that could be computed, Wood et al. obtained a nonsignficant mean value of −0.01. The mean effect size for the 18 representative samples was −0.05, again indicating gender similarities in well-being. Effect size varied with type of measure, but all were close to zero.

To assess the effect of marital status, Wood et al. used the percentage of the respondents in the sample who were married as a predictor variable in a regression-type analysis. The effect was significant and indicated that studies with a higher percentage of married persons obtained larger effect sizes favoring women. The validation analysis yielded the same result, and the general finding held for each type of dependent measure. Additional analyses revealed that marriage is associated with enhanced well-being for both men and women, but that this difference tends to be greater for women. Wood et al. accounted for this result within the framework of social role theory. They argued that women's social role is associated with greater emotional sensitivity, expressiveness, and skillfulness and that marriage and family life provide women with greater opportunities to fulfill their gender role of "emotional specialist."

Kling, Hyde, Showers, and Buswell (1999) used a developmental approach in their meta-analysis of studies of gender differences in self-esteem, based on the assertion of prominent authors such as Pipher (1994) that girls' self-esteem takes a nosedive at the beginning of adolescence. Kling et al. found that the magnitude of the gender difference did grow larger from childhood to adolescence: in childhood (ages 7–10), d was 0.16; for early adolescence (ages 11–14), it was 0.23; and for the high school years (ages 15–18), 0.33. However, the gender difference did not suddenly become large in early adolescence and, even in high school, the difference was still not large. Moreover, the gender difference was smaller in older samples; for example, for ages 23 to 59, *d* was 0.10.

Kling and colleagues also analyzed the magnitude of gender differences as a function of ethnicity. For whites, *d* was 0.20, whereas for blacks, it was −0.04. Therefore, the gender difference in self-esteem, which is small among whites, is nonexistent among blacks, calling into question supposedly well-known psychological "facts" that are based on white samples. In this meta-analysis, too few studies reporting data on self-esteem in other ethnic groups were available for analysis.

To assess gender differences in childhood depression Twenge and Nolen-Hoeksema (2002) examined 310 studies that assessed depression with the Childhood Depression Inventory (CDI; Kovacs, 1985, 1992) among children between the ages of 8 and 16. Moderator analyses suggested that the overall effect size of 0.02 was significantly moderated by age. Specifically, there were no gender differences in CDI scores between the ages of 8 and 12 ($d = -0.04$ with 86 studies); however, girls scored higher on the CDI starting at age 13 ($d = 0.08$). At ages 14 and 15, the differences reached 0.22 and remained significantly different at age 16 ($d = 0.18$). The authors further demonstrated that when all samples for ages 13 to 16 were combined (49 studies) the overall effect size was 0.16, suggesting that while gender differences in depression were not apparent during childhood, they were significant during adolescence. The finding that boys' depression remains relatively stable between the ages of 8 and 16, whereas girls' depression begins to steadily increase after age 12 supports the notion that gender differences in depression emerge during adolescence.

META-ANALYSIS AND GENDER DIFFERENCES IN MOTOR ACTIVITY LEVEL AND MOTOR PERFORMANCE

Motor activity level has been defined as an "individual's customary level of energy expenditure through movement" (Eaton & Enns, 1986, p. 19). It is conceived of as an important component of temperament and can even be measured prenatally (e.g., Robertson, Dierker, Sorokin, & Rosen, 1982). The single meta-analysis performed to date on gender differences in motor activity level was done by Eaton and Enns (1986). They evaluated 127 independent effect sizes taken from 90 different research reports and examined the effects of developmental factors, situational factors, measurement factors, and investigator factors on the size of *d*. It is important to note that in 90 percent of the studies included in the analysis, the mean age of the sample was 15 years or less. Consequently, the results of the Eaton and Enns work are not necessarily applicable to older persons.

Across all studies, Eaton and Enns obtained an average effect size of 0.49, indicating a higher activity level for males. However, they found small and significant correlations between *d* and subjects' age ($r = 0.26$), the restrictiveness of the setting where the measurements were taken (e.g., playground versus classroom; $r = -0.22$), and the inclusiveness of the measurement instrument used (e.g., a low-inclusive instrument would be one that measured arm movements, whereas a high-inclusive instrument would be one that measured whole-body movements and general activity level; $r = -0.28$). A multiple regression analysis indicated that larger effect sizes were found in studies of older (i.e., preadolescent and adolescent) youths whose behavior was assessed in nonstressful, unrestrictive settings and in the presence of peers.

Thomas and French (1985) performed a meta-analysis on 64 studies of gender differences in motor performance, from which they computed 702 effect sizes. Of the 20 motor tasks included in the analysis, 12 were found to yield age-related effect size curves. For eight of these tasks (balancing, catching, grip strength, pursuit rotor, shuttle run, tapping, throw velocity, and vertical jump), the relationship between age and *d* was a positive linear one; for the remaining four (dash, long jump, sit-ups, and throw distance), the relationship was a quadratic one (U-shaped). The eight tasks that did not yield age-related gender differences were agility, anticipation timing, arm hang, fine eye-motor, flexibility, reaction time, throw accuracy, and wall volley. For 18 of the 20 tasks, the mean effect size across studies was positive, indicating better performance by males. Most of these values ranged between 0.01 and 0.66, with the mean effect sizes for throw velocity and throw distance being much larger (2.18 and 1.98, respectively). Only the fine eye-motor and flexibility tasks yielded negative mean effect sizes (−0.21 and −0.29, respectively), indicating better female performance.

Thomas and French concluded that the data "do not support the notion of uniform development of gender differences in motor performance across childhood and adolescence" (p. 273). They argued that before puberty, the performance differences between girls and boys are typically small to moderate (*d* values of 0.20–0.50), meaning that many girls are outperforming boys. They further argued that these prepubertal differences are most likely the result of environmental factors (e.g., parent and teacher expectations and encouragement, practice opportunities, and so on) and not biological ones. Then, at puberty, the greater increase in boys' size and muscle development—combined with the continued and perhaps intensified environmental influences—results in a greater gender gap in motor performance that continues through adolescence. Evidence that female Olympic athletes have continued to close the gender-related performance gap on both the 100-meter dash and 100-meter freestyle swimming events suggests that gender differences in motor performance are highly responsive to environmental forces such as training and need not persist into adulthood (Linn & Hyde, 1989).

CONCLUSION

We believe that meta-analysis is a useful tool that can advance the study of gender differences and similarities, for several reasons:

1. Meta-analysis indicates not only whether there is a significant gender difference but also how large the difference is. Therefore, it can be used to determine which psychological gender differences are large and which are small or trivial.
2. Meta-analysis represents an advance over years of psychological doctrine stating that one could never accept the null hypothesis. We believe that some effect sizes obtained through meta-analysis are so small that the null hypothesis of no gender difference can be accepted (Hyde & Linn, 1988). We recommend that any effect size less than 0.10 be interpreted as no difference. This in turn will allow researchers to lay to rest some persisting rumors of psychological gender differences that are simply unfounded.
3. One of the most important trends in gender research today is the investigation of gender x situation interactions; meta-analysis permits some powerful analyses of this sort. Eagly's report of the situations that promote different patterns of gender differences in helping behaviors is an excellent example (Eagly & Crowley, 1986).
4. Meta-analysis can be a powerful tool to analyze issues other than gender differences. Examples include analyses of the role of androgyny and masculinity or ethnicity in women's psychological well-being (e.g., Bassoff & Glass, 1982; Grabe & Hyde, 2006). As a further example, feminist psychologists are increasingly interested in investigating the joint effects of gender and ethnicity; meta-analysis can be used here as well. For example, Hyde, Fennema, and Lamon (1990) examined gender differences in math performance as a function of ethnicity and found mean d values of −0.02, 0.00, −0.09, 0.13, for blacks, Latinos, Asian Americans, and whites, respectively.
5. Meta-analyses can be theory grounded and can be used to test theories of gender. Good examples are Eagly's application of social role theory in predicting patterns of gender differences in aggression and in helping behaviors (Eagly & Crowley, 1986; Eagly & Steffen, 1986).
6. Finally, meta-analysis can be used to test the Gender Similarities Hypothesis, which stands in stark contrast to the differences model that holds that men and women, and boys and girls, are vastly difference psychologically. The Gender Similarities Hypothesis states, instead, that males and females are alike on most—but not all—psychological variables (Hyde, 2005).

REFERENCES

AnnOnline (2007). Biography: Deborah Tannen. Retrieved February 14, 2007, from http://www.annonline.com/interviews/990310/biography.html.

Archer, J. (2004). Sex differences in aggression in real-world setting: A meta-analytic review. *Review of General Psychology, 8,* 291–322.

Bassoff, B. S., & Glass, G. V. (1982). The relationship between sex roles and mental health: A meta-analysis of 26 studies. *Counseling Psychologist, 10,* 105–112.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155–162.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Eagly, A. H. (1986). Some meta-analytic approaches to examining the validity of gender difference research. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 159–177). Baltimore: Johns Hopkins University Press.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100,* 283–308.

Eagly, A. H., Johannesen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin, 129,* 569–591.

Eagly, A. H. & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin, 108,* 233–256.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and effectiveness of leaders: A meta-analysis. *Psychological Bulletin, 117,* 125–145.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111,* 3–22.

Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100,* 309–330.

Eaton, W. O. & Enns, L. R. (1986). Sex differences in human motor activity level. *Psychological Bulletin, 100,* 19–28.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Goldstein, A. G. & Jeffords, J. (1981). Status and touching behavior. *Bulletin of Psychonomic Society, 17,* 79–81.

Grabe, S., & Hyde, J. S. (2006). Ethnicity and body dissatisfaction among women in the United States: A meta-analysis. *Psychological Bulletin, 132,* 622–640.

Gray, J. (1992). *Men are from Mars, women are from Venus.* New York: HarperCollins.

Gray, J. (2007). *John Gray, Ph.D. is the best-selling relationship author of all time.* Retrieved February 14, 2007, from http://www.marsvenus.com/john-gray.php.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style.* Baltimore: Johns Hopkins University Press.

Hall, J. A., & Halberstadt, A. G. (1986). Smiling and gazing. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 136–158). Baltimore: Johns Hopkins University Press.

Hare-Mustin, R. T., & Marecek, J. (1988). The meaning of difference: Gender, theory, postmodernism, and psychology. *American Psychologist, 43,* 455–464.

Hedges, L. V. (1982a). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7,* 119–137.

Hedges, L. V. (1982b). Fitting continuous models to effect size data. *Journal of Educational Statistics, 7,* 245–270.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Henley, N. M. (1977). *Body politics: Power, sex, and nonverbal communication.* Englewood Cliffs, NJ: Prentice-Hall.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Hyde, J. S. (1984). How large are gender differences in aggression? A developmental meta-analysis. *Developmental Psychology, 20,* 722–736.

Hyde, J. S. (1985). *Half the human experience: The psychology of women* (3rd ed.). Lexington, MA: D. C. Heath.

Hyde, J. S. (1986). Introduction: Meta-analysis and the psychology of gender. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 1–13). Baltimore: Johns Hopkins University Press.

Hyde, J. S. (1990). Review essay: Meta-analysis and the psychology of gender differences. *Signs, 16,* 55–73.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592.

Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107,* 139–155.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14,* 299–324.

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

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

Kling, K. C., Hyde, J. S., Showers, C., & Buswell, B. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin, 125,* 470–500.

Kovacs, M. (1985). The Children's Depression Inventory (CDI). *Psychopharmacology Bulletin, 21,* 995–998.

Kovacs, M. (1992). *Children's Depression Inventory Manual.* North Tonawanda, NY: Multi-Health Systems.

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin, 129,* 305–334.

Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher, 18,* 17–27.

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

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Pipher, M. (1994). *Reviving Ophelia: Saving the selves of adolescent girls.* New York: Ballantine.

Robertson, S. S., Dierker, L. J., Sorokin, Y., & Rosen, M. G. (1982). Human fetal movement: Spontaneous oscillations near one cycle per minute. *Science, 218,* 1327–1330.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Newbury Park, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169.

Shields, S. A. (1975). Functionalism, Darwinism, and the psychology of women: A study in social myth. *American Psychologist, 30,* 739–754.

Spence, J. T., Helmreich, R., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity-femininity. *JSAS Catalog of Selected Documents in Psychology, 4,* 43 (Ms. No. 617).

Stier, D. S., & Hall, J. A. (1984). Gender differences in touch: An empirical and theoretical review. *Journal of Personality and Social Psychology, 47,* 440–459.

Tannen, D. (1991). *You just don't understand: Women and men in conversation.* New York: Ballantine.

Taylor, M. C., & Hall, J. A. (1982). Psychological androgyny: Theories, methods, and conclusions. *Psychological Bulletin, 92,* 347–366.

Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance: A meta-analysis. *Psychological Bulletin, 98,* 260–282.

Twenge, J. M., & Nolen-Hoeksema, S. (2002). Age, gender, race, socioeconomic status, and birth cohort differences on the Children's Depression Inventory: A meta-analysis. *Journal of Abnormal Psychology, 111,* 578–588.

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

Weitzman, L. J. (1985). *The divorce revolution: The unexpected social and economic consequences for women and children in America.* New York: Free Press.

Wood, W. (1987). Meta-analytic review of sex differences in group performance. *Psychological Bulletin, 102,* 53–71.

Wood, W., Rhodes, N., & Whelan, M. (1989). Sex differences in positive well-being: A consideration of emotional style and marital status. *Psychological Bulletin, 106,* 249–264.