Evaluating Gender Similarities and Differences Using Metasynthesis

Ethan Zell   *University of North Carolina at Greensboro*
Zlatan Krizan   *Iowa State University*
Sabrina R. Teeter   *Western Carolina University*

Despite the common lay assumption that males and females are profoundly different, Hyde (2005) used data from 46 meta-analyses to demonstrate that males and females are highly similar. Nonetheless, the gender similarities hypothesis has remained controversial. Since Hyde's provocative report, there has been an explosion of meta-analytic interest in psychological gender differences. We utilized this enormous collection of 106 meta-analyses and 386 individual meta-analytic effects to reevaluate the gender similarities hypothesis. Furthermore, we employed a novel data-analytic approach called metasynthesis (Zell & Krizan, 2014) to estimate the average difference between males and females and to explore moderators of gender differences. The average, absolute difference between males and females across domains was relatively small (d = 0.21, SD = 0.14), with the majority of effects being either small (46%) or very small (39%). Magnitude of differences fluctuated somewhat as a function of the psychological domain (e.g., cognitive variables, social and personality variables, well-being), but remained largely constant across age, culture, and generations. These findings provide compelling support for the gender similarities hypothesis, but also underscore conditions under which gender differences are most pronounced.

Keywords: gender, gender differences, gender similarities, meta-analysis

Supplemental materials: http://dx.doi.org/http://dx.doi.org/10.1037/a0038208.supp

All of us are aware of gender stereotypes that espouse profound differences between males and females. Presumably, males are from Mars and females are from Venus, males are tough and females are tender, males are competitive and females are cooperative, males are dominant and females are submissive, males are stoic and females are emotional, males are quiet and females are talkative, males are mathematical and females are verbal, and so on. The assumption of these and many other prevalent gender stereotypes is that males and females are vastly different in their personality, abilities, interests, attitudes, and behavioral tendencies. This assumption is referred to as the *gender differences hypothesis* (Gray, 1992; Tannen, 1991). But is it correct? Are males and females really that different? The current report aggre-gates data from 106 independent meta-analyses with over 12 million participants to test the gender differences hypothesis and to explore moderators of gender differences. In doing so, we provide the most comprehensive analysis of gender differences to date.

On the one hand, it is not surprising that people assume large differences between males and females. Along with age and race, gender is perhaps the most salient category that guides social perception (Macrae & Bodenhausen, 2000). Perceivers rapidly categorize a person's gender and immediately draw inferences about them using gender as a cue (Ito & Urland, 2003). Moreover, there are typically obvious anatomical and biological differences between males and females (i.e., biological sex). People may assume that males and females differ psychologically to a similar extent that they differ physically. Finally, people are repeatedly exposed to cultural stereotypes on supposed gender differences, starting in childhood (e.g., Browne, 1998). This barrage of cultural messages may create the illusion that such stereotypes are correct and could guide how people perceive and interpret the world around them.

On the other hand, behavioral scientists have been conducting rigorous tests of the gender differences hypothesis for several decades, and evidence for its core assumption has been sparse. Researchers have increasingly used meta-analysis to test the gender differences hypothesis by examining the overall, average difference between males and females across numerous studies in a given domain. For example, highly cited meta-analyses have examined gender differences in math performance (Hyde, Fennema, & Lamon, 1990), self-esteem (Kling, Hyde, Showers, & Buswell, 1999), personality (Feingold, 1994), and aggression (Archer, 2000). In each of these specific cases, gender differences were found to be relatively small. Based on emerging meta-analytic findings, theorists have proposed an alternative perspective, known as the *gender similarities*

Ethan Zell, Department of Psychology, University of North Carolina at Greensboro; Zlatan Krizan, Department of Psychology, Iowa State University; Sabrina R. Teeter, Department of Psychology, Western Carolina University.

We thank Yanna Weisberg for comments on a previous version of this article.

Correspondence concerning this article should be addressed to Ethan Zell, Department of Psychology, University of North Carolina at Greensboro, P.O. Box 26170, Greensboro, NC 27402. E-mail: e_zell@uncg.edu

Ethan Zell

hypothesis (Hyde, 2005). That is, males and females may be similar on most (but not all) psychological dimensions, and when differences do arise, they should typically be small in magnitude.

Estimating Gender Differences

Meta-analysis has proven to be a powerful tool to assess gender differences in specific domains (see Hyde, 2014). However, the fundamental question of how different males and females are across domains remains largely unresolved. Indeed, prominent cultural stereotypes not only pertain to supposed differences in specific domains (e.g., math ability or leadership skill), but they also pertain to global differences between males and females across domains (e.g., males are from Mars, females are from Venus). Although domain-specific findings may help deflate bogus stereotypes in a given domain, other stereotypes as well as the general impression that males and females are fundamentally different may remain. For example, when confronted with findings demonstrating that females perform just as well as males on math tests (Else-Quest, Hyde, & Linn, 2010; Lindberg, Hyde, Petersen, & Linn, 2010), people may revise or even abandon stereotypes about gender differences in math. However, they will most likely retain gender stereotypes about other domains, as well as the more basic assertion that the genders differ in profound ways. Thus, to address the more basic question of how males and females differ across domains, researchers need to go beyond meta-analyses in specific domains.

As meta-analyses on gender differences have been accumulating, researchers have begun to aggregate meta-analytic findings to derive more global estimates of the difference between males and females. Richard, Bond, and Stokes-Zoota (2003) were the first group of scholars to examine gender differences across domains by aggregating meta-analytic findings using metasynthesis (i.e., second-order meta-analysis; see Johnson, Scott-Sheldon, & Carey, 2010; Zell & Krizan, 2014). Specifically, Richard and colleagues (2003) aggregated data from 34 meta-analyses examining topics related to social and personality psychology (e.g., attribution, relationships, and nonverbal communication). The overall, absolute difference between males and females was found to be a *d* of .24, which would be classified as a small effect using conventional standards (Cohen, 1988).

Going further, Hyde (2005) collected data from 46 "major" meta-analyses examining gender differences across several psychological domains, including social and personality variables, cognitive variables, and psychological well-being. Although procedures were not used to quantitatively aggregate the findings, 48% of the meta-analytic effects were small and 30% were very small or close to zero. This pattern of results led Hyde to conclude that meta-analytic findings provide stronger support for the gender similarities hypothesis than the gender differences hypothesis. Hyde's article has become a classic in the psychological canon, as evidenced by the fact that it has been cited 1,430 times according to Google Scholar and 620 times according to Scopus (as of August 2014). However, the article has also been regarded as highly controversial, as the basic question of whether existing data better support the gender similarities hypothesis or the gender differences hypothesis remains hotly contested (see Carothers & Reis, 2013; Eagly & Wood, 2013; Stewart-Williams & Thomas, 2013). Along these lines, skeptics have noted that Hyde's (2005) findings were limited by an analysis of only about 40 psychological domains, left out several key topics that had not been meta-analyzed at that point, and broadly conflicted with large gender differences in mating preferences identified by evolutionary psychologists (e.g., Buss, 1989, 2013; Schmitt et al., 2012).

The Current Report

Since the publication of Hyde's (2005) analysis, there has been an explosion of meta-analytic interest in gender differences. New meta-analyses have examined differences between males and females in domains such as cooperation (Balliet, Li, Macfarlan, & Van Vugt, 2011), impulsivity (Cross, Copping, & Campbell, 2011), self-conscious emotions (Else-Quest, Higgins, Allison, & Morton, 2012), language use (Leaper & Ayres, 2007), and interests (Su, Rounds, & Armstrong, 2009). In addition, new meta-analyses have been conducted on topics previously studied (e.g., sexuality), using more sophisticated statistical techniques and a larger pool of data (Petersen & Hyde, 2010). With this enormous and highly diverse collection of over 100 meta-analyses (described in more detail in the Metasynthesis Method section), the psychological literature has crossed the threshold whereby new insights about overall gender differences can now be made. The current study utilizes this vast collection of meta-analytic findings to derive the most comprehensive test of the gender similarities hypothesis to date. Further, the current report

Zlatan Krizan

examined several potential moderators of gender differences to ascertain their generalizability, including age, culture, and time period, as well as the psychological domain.

Metasynthesis Method

Before getting to our specific procedures and findings, it is important to note the unique strengths and weaknesses of metasynthesis in this context (see Cooper & Koenka, 2012). On the positive side, metasynthesis affords the opportunity to assess the global question of gender differences across domains. Although this question could also be tested using meta-analysis, a meta-analytic approach would be highly impractical due to the enormous amount of primary studies (over 20,000). By using metasynthesis, we were able to incorporate data from these studies without having to track down and code all of the original articles. Further, because metasynthesis often utilizes extremely large pools of data and participants, it affords highly reliable conclusions. At a time when psychological science has been criticized for underpowered studies that sometimes fail to replicate (Simmons, Nelson, & Simonsohn, 2011), metasynthesis is a useful retort as it affords enormous power and confident conclusions.

On the negative side, given its broader approach metasynthesis cannot be used to evaluate more finite microlevel hypotheses that are often of interest to psychological scientists (see Eagly & Wood, 2013). Additionally, because individual meta-analyses often use different statistical models to aggregate primary data (e.g., fixed-effect, random-effects), it is inappropriate to formally aggregate meta-analytic findings to conduct tests of statistical significance. However, given that metasynthesis typically has extremely high power, tests of statistical significance would not be very informative, as even minute differences

would yield a statistically significant outcome. A final concern of metasynthesis is the overlapping of samples. Specifically, there are occasions where a single study can be included in multiple, related meta-analyses. Therefore, if metasynthesis incorporated all of these meta-analyses, the individual study would be counted more than once violating assumptions regarding independence of observations. However, when meta-analyses share large amounts of data, researchers can simply select one of these meta-analyses for inclusion in the metasynthesis (e.g., the more recent one).

Identifying Meta-Analyses

We obtained meta-analyses on psychological gender differences, defined as gender differences in mind and behavior (American Psychological Association, 2014) by using two approaches. First, we scanned existing reviews of the gender differences literature for relevant articles (e.g., Eagly & Wood, 2013; Hyde, 2005, 2014; Richard et al., 2003). Second, we searched for meta-analyses using databases such as PsychInfo, Google Scholar, and Dissertations Abstracts International during April of 2014. The search terms "sex differences" and "meta-analysis" were entered simultaneously. We also conducted a search using the terms "gender differences" and "meta-analysis." Together, these searches yielded over 500 relevant papers.

Many of the articles identified were immediately excluded on the basis that they were not meta-analyses or did not examine gender differences. Studies that dealt with outcomes that were not psychological in nature were excluded (e.g., medical outcomes, biological or anatomical differences; Thomas & French, 1985). Articles that examined differences as a function of perceived gender roles as opposed to gender per se were excluded. Gender roles involve the degree to which people adopt stereotypically masculine versus feminine traits, behaviors, and interests, rather than their gender identity (i.e., whether they identify as male or female; see Reilly & Neumann, 2013). Studies were excluded if they examined gender only as a function of other factors (i.e., did not report a direct gender comparison). Articles were also excluded if they did not report an overall, mean difference between males and females using a standard metric (e.g., Cohen's d, Hedges' g, or Pearson's r). Finally, we excluded articles that focused on specialized, nonrepresentative populations (e.g., people with major psychoses or prisoners). However, we did not exclude articles on the basis of age or cultural group, so that the potential influence of these variables could be tested in moderation analyses.

Next, we evaluated meta-analyses to address the possibility of overlapping samples. First, meta-analyses were excluded if they were replaced by newer meta-analyses on the same topic that incorporated additional studies or if they analyzed only a subset of studies of an existing meta-analysis. Second, we compared the reference sections of related papers to quantify the degree of sample overlap. Although there were several cases of sample overlap, in no instance did the degree of overlap exceed 25%. This means that for each meta-analysis we examined, at least 75% of its samples were unique

Sabrina R. Teeter

(i.e., unshared with other meta-analyses). Therefore, the meta-analyses incorporated into our model were mostly independent, in that they largely contained unique data that was not shared by other meta-analyses.

After removing articles that did not fit the inclusion criteria, our search yielded a total of 106 meta-analyses examining psychological gender differences (see Supplemental Table 1). The final set of articles covered an enormous range of topics such as interruption, risk taking, helping behavior, leadership styles, body image, intelligence, occupational stress, jealousy, and morality, among other topics (see Supplemental References). Across meta-analyses, the total number of effects was 21,174. Several articles did not provide sufficient information to calculate sample size ($m = 10$). However, among the remaining papers, the total sample size was 12,238,667 participants.

Study Treatment and Analyses

Meta-analytic effects were obtained from each article as an estimate of effect size in a given topic area (e.g., math ability, aggression). We focus on the *absolute value* of each meta-analytic effect because our purpose was to assess the magnitude of the difference between males and females regardless of the particular direction of difference (for a similar approach, see Hyde, 2005; Richard et al., 2003). Further, directional tests are more appropriate in traditional meta-analyses where the goal is to assess whether males score higher or lower than females in a given topic area. Finally, the use of absolute values helps prevent misleadingly low estimates of gender differences. If males score higher than females in one topic area (+0.5) and females score higher than males in another topic area (−0.5), the average of these effects would be 0, implying no overall gender difference. However, averaging absolute effects

would yield an overall estimate of 0.5, which we argue better represents the magnitude of global gender differences.

Where possible, we used the uncorrected, unweighted difference between males and females as an estimate of effect size (d). However, some meta-analyses only reported weighted or corrected values (e.g., Hedges' g). Effect sizes that were reported in the r metric were converted to Cohen's d. The model we employed used an unweighted average of the individual meta-analytic effects as an estimation of the population effect. In traditional meta-analyses, unweighted averages are robust and tend to outperform averages that weight by study sample size (Bonett, 2009; Krizan, 2010; Shuster, 2010). We focus primarily on aggregate estimates and their ranges, without formal computation of relevant confidence intervals. Our descriptive focus is necessitated by the fact it is inappropriate to calculate confidence intervals by formally aggregating estimates that derive separately from fixed-effect and random-effects models (Hedges & Vevea, 1998). In addition, many studies did not provide information essential for such computations (e.g., confidence intervals or standard errors).

We use Cohen's (1988) conventions to interpret the magnitude of gender differences (d). Cohen argued that the cutoff points of .2, .5, and .8 reflect small, medium, and large differences, respectively. Further, Hyde (2005) argued that effect sizes of .10 or below should be considered very small. When interpreting effects, we also present the percentage of overlap between male and female distributions. Specifically, Cohen (1988) argued that effect sizes .2, .5, and .8 correspond to distributions that overlap by approximately 85%, 67%, and 52%, respectively. Distributions with a large amount of overlap can be viewed as highly similar. When describing our results, we use the following symbols: d (average absolute difference between males and females in standard deviation units), k (number of effects), and m (number of meta-analyses).

Magnitude of Gender Differences

Preliminary Analyses

Most meta-analyses provided only a single effect ($m = 61$); however, other meta-analyses provided multiple effects ($m = 45$), with some reporting over 30 individual meta-analytic tests in their report. For example, Petersen and Hyde (2010) conducted separate meta-analytic tests of gender differences in 30 sexual attitudes and behaviors (e.g., condom use, premarital sex, same-gender sex). In total, we obtained 386 meta-analytic effects, an amount that is several times larger than the 124 obtained by Hyde (2005). As can be seen in Table 1, the majority of these effects were either very small (39.4%) or small (46.1%); relatively few effects were medium (11.9%), large (1.8%), or very large in size (0.8%).

Further, we scanned meta-analyses for information regarding the homogeneity of effects (e.g., I^2, T, T^2, Q). Effects that are largely consistent (inconsistent) in size and direction are considered homogeneous (heterogeneous). Direct homogeneity tests or statistics were not reported for

Table 1
Meta-Analytic Effect Sizes (m = 386) by Range of Magnitude

Effect sizes	Effect size range				
	0–0.10 Very small	0.11–0.35 Small	0.36–0.65 Medium	0.66–1.00 Large	>1.00 Very large
Number	152	178	46	7	3
% of total	39.4	46.1	11.9	1.8	0.8

83 effects. Of the remaining effects, 92 (30.4%) were homogeneous and 211 (69.6%) were heterogeneous. Therefore, although most of the meta-analytic effects obtained in this report were small or very small, individual-study effects within these meta-analyses were typically heterogeneous, suggesting that gender differences often vary by context.

Primary Model

Our primary model averaged meta-analytic effects within papers when more than one was provided. Several of the papers that derived multiple effects utilized some or all of the same participants in multiple meta-analytic effects. Incorporating each of these effects into the model would violate independence of observations. Thus, we incorporate only a single effect size from each individual meta-analysis in our primary model as well as the moderation tests reported below. Specifically, for papers that provided multiple effects in a given topic area (e.g., interests), we averaged the absolute value of each effect as an estimate of gender differences in that topic area.

When considering the entire collection of 106 meta-analyses, the average absolute difference between males and females across topic areas was a *d* of .21, typically regarded as a small effect (see Figure 1). There was some dispersion in the individual effects (*SD* = .14), with the smallest effect being a *d* of .02 and the largest being a *d* of .73 (see Table 2). Only one of the meta-analytic averages would be characterized as large in size,

and 81% of the effects were between .01 and .30. Put simply, more than three quarters of the observed gender differences reflected almost 80% overlap across distributions of males and females. This finding provides strong support for the gender similarities hypothesis. That is, although theorists and laypersons have long assumed that males and females are profoundly different (e.g., Gray, 1992; Tannen, 1991), our findings suggest that these assertions are likely inflated. We conducted a supplemental analysis that weighted meta-analytic effects by the number of effect sizes. This alternative, weighted model yielded an effect size (.19) that was highly similar to that obtained by our primary model (.21). Therefore, we retained the use of an unweighted model in subsequent analyses.

Although gender differences were typically small, there were several effects that were moderate to large in size. The 10 largest gender differences in the psychological literature are presented in Table 3. According to these findings, males score higher than females on measures of masculinity, mental rotation ability, importance of physical attractiveness in mate selection, and aggression. Females score higher than males on measures of reactivity to painful (noxious) stimuli, peer-attachment, and interest in people as opposed to things.

Figure 1
Visual Depiction of Two Distributions That Differ by an Effect Size (d) of 0.21 (see Magnusson, 2014)

Table 2
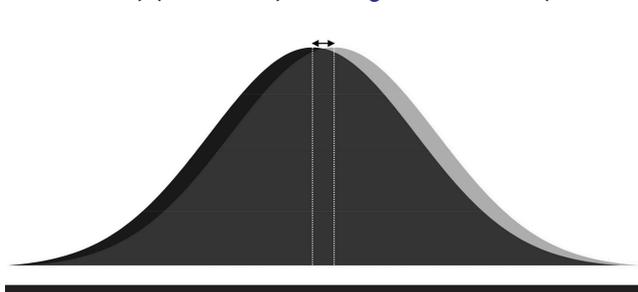*Stem-and-Leaf Display of 106 Meta-Analytic Averages*

Stem	Leaf
.7	3
.6	
.5	1 3 6 7
.4	0 1 5 9
.3	0 0 0 1 4 4 4 5 6 7 7 8 9
.2	0 0 0 0 0 0 1 1 1 2 2 3 3 3 4 4 4 5 5 5 5 6 6 6 7 7 8 8 8
.1	0 0 0 0 0 0 0 1 1 1 2 2 2 2 2 3 3 4 4 4 4 5 6 6 6 6 6 7 8 8 8 9 9 9
.0	2 2 2 2 3 3 4 4 5 5 5 5 5 7 7 8 8 8 9

Note. Values represent absolute effects sizes (Cohen's *d*).

Table 3
Ten Largest Gender Differences

Topic	d	k	Result	Reference
Masculine vs. feminine traits	.73	59	M > F	Twenge, 1997
Mental rotation ability	.57	70	M > F	Maeda & Yoon, 2013
Noxious stimulation	.56	26	F > M	Riley et al., 1998
Importance of beauty in mates	.53	28	M > F	Feingold, 1990
Peer attachment	.51	43	F > M	Gorrese & Ruggieri, 2012
Interest in people vs. things	.49	745	F > M	Su et al., 2009
Aggression	.45	197	M > F	Knight et al., 2002
Film-induced fear	.41	95	F > M	Peck, 2000
Confidence in physical abilities	.40	46	M > F	Lirgg, 1991
Same-sex group performance	.39	64	M > F	Wood, 1987

Note. d = average absolute effect size; k = number of effects; M > F indicates that males scored higher than females. Full references are provided in the Supplemental Materials.

Moderators of Gender Differences

Theoretical Moderators

Psychological domain. Gender differences research is broadly clustered into three psychological domains: cognitive variables, social and personality variables, and well-being (see Hyde, 2014). We examined whether the magnitude of gender differences varies across these fundamental domains of psychological functioning. Cognitive variables were defined as mental processes such as attention, memory, and problem solving, including variables such as math performance, spatial performance, and verbal skills. Social and personality variables were defined as constructs related to the study of individual differences or social behavior and they included variables such as temperament, interests, aggression, interpersonal communication, helping, sexuality, and leadership. Finally, well-being was defined as any variable related to positive or negative mental health. This included variables such as depression, rumination, and self-esteem. Two raters coded the domain of study for each meta-analysis ($\kappa = .88$), and differences between raters were resolved through discussion.

There was some fluctuation in effect sizes by domain (see Table 4 for descriptive statistics, including variability measures). Gender differences were comparable when examining research on cognitive variables (.22) and social and personality variables (.22), but were somewhat smaller when examining research on psychological well-being (.14). Further, the range of differences was more condensed for research on well-being. These findings show that gender differences are relatively small regardless of domain, but that differences in well-being are particularly small. Nonetheless, it is important to note that there were fewer meta-analyses on psychological well-being ($m = 11$) than cognitive variables ($m = 30$) or social and personality variables ($m = 65$), and thus, it is possible that the effect size in this domain will increase as more meta-analyses are conducted.

Separate stem-and-leaf plots for each psychological domain show that the distribution of effect sizes was similar across domains, albeit more condensed for studies on well-being (see Supplemental Tables 2–4). Further, the effects of other moderators that we discuss below (age, culture, and time period) were largely comparable when examining effects within each psychological domain (see Supplemental Tables 5–7).

Age. Little is known about the degree to which gender differences change over the life span. The gender differences hypothesis presumes that there are large differences between males and females at all ages, but it is also possible that gender differences are small at all ages or fluctuate as a function of age. Along these lines, meta-

Table 4
Theoretical Moderators of Gender Differences

Moderator	m	k	d	SD	Range
Research domain					
Cognitive	30	3,611	.22	.13	.05–.57
Social/personality	65	15,590	.22	.14	.02–.73
Well-being	11	1,973	.14	.09	.02–.28
Age					
Child/adolescent	8	1,914	.17	.10	.02–.34
Adult	32	2,928	.18	.15	.02–.73
Mixed	66	16,332	.23	.13	.02–.57
Culture					
Euro-American	46	10,508	.19	.14	.02–.73
Multicultural	47	8,902	.22	.11	.03–.51
Time period					
1980s	10	841	.15	.11	.02–.39
1990s	34	4,759	.23	.16	.02–.73
2000s	32	8,311	.22	.12	.02–.49
2010s	30	7,263	.19	.13	.03–.57

Note. m = number of meta-analyses; k = number of effects; d = average absolute effect size; SD = standard deviation of the absolute effect size.

analyses in specific topic areas have shown somewhat larger gender differences in adolescents and young adults relative to children with regards to self-esteem (Kling et al., 1999), depression (Twenge & Nolen-Hoeksema, 2002), and mathematics performance (Hyde et al., 1990). However, these findings do not indicate the degree to which global differences between males and females change across the life span.

To further explore this issue, we scanned meta-analyses for data relevant to age. Most meta-analyses utilized a mix of child, adolescent, and adult participants ($m = 66$). However, some meta-analyses examined only adult participants ($m = 32$), and others examined only child and/or adolescent participants, defined as participants age 18 years and below ($m = 8$). Effect sizes were small in size regardless of the age of the samples utilized in meta-analyses. Specifically, meta-analyses that examined gender differences in children and/or adolescents yielded an average effect size of .17. Studies examining adults yielded a similar average effect size of .18. Therefore, global differences between males and females are relatively small and appear to remain constant from childhood to adulthood. Moreover, studies examining children and/or adolescents covered a similar range of topics than studies examining adults (Supplemental Tables 5–7), which suggests that age was likely not confounded with psychological domain.

Culture. We also tested whether global differences between males and females vary across cultural groups. Some meta-analyses did not provide sufficient information to determine the country or culture of the participants studied ($m = 12$), and one meta-analysis included only participants from Turkey (Aydin, Sarier, & Uysal, 2011). Of the remaining meta-analyses, samples consisted of only European American participants ($m = 46$) or participants from at least two different cultural groups (e.g., European American, East Asian, African, Middle Eastern; $m = 47$). Meta-analyses that examined only European Americans yielded an effect size of .19. Similarly, meta-analyses that examined multiple cultures yielded an effect size of .22. Thus, differences between males and females across domains are relatively small and appear to remain constant when comparing European American samples to more diverse samples. However, few of the multicultural meta-analyses indicated the proportion of effects from European American samples versus other samples. It is possible that these papers contained effects derived predominantly from European American samples.

Time period. It has been argued that gender differences in specific domains such as personality and competence may have declined over the last several decades, as women's roles have expanded to become more comparable to that of men (Eagly & Wood, 2013; Hyde, 2005). We examined whether global differences between males and females are increasing or decreasing over time. The average difference between males and females was small, whether looking at meta-analyses published in the 1980s (.15), 1990s (.23), 2000s (.22), or 2010s (.19). Further, the year of publication was not substantially correlated with effect size magnitude, $r = .03$, $p = .80$. These findings

suggest that differences between males and females across domains have remained largely constant over the last several decades. It should be noted that our analysis is not a strong test of differences over time, as newer meta-analyses typically incorporate both newer and older data.

Methodological and Statistical Moderators

Quality. We examined whether meta-analyses that used more sophisticated methodologies yielded different estimates of gender differences than those that used less sophisticated methodologies (see Cooper & Koenka, 2012). Along these lines, we created a five-item checklist to evaluate the quality of each meta-analysis, by selecting the five most relevant items from a larger checklist developed to evaluate meta-analyses in medicine (Higgins et al., 2013). Specifically, two coders evaluated meta-analyses in terms of the search for articles, inclusion/exclusion of articles, coding of articles, assessment of heterogeneity, and data synthesis methodology ($\kappa s > .81$).

Most meta-analyses scored favorably, with the average meta-analysis receiving a score of 4 out of 5 on the quality checklist ($M = 3.91$, $SD = 1.04$). Although there was some fluctuation in effect sizes as a function of methodological quality, effect sizes were small regardless of quality (see Table 5). Further, a correlation analysis showed that effect sizes were somewhat smaller in high quality than low quality meta-analyses, but this effect was not statistically significant, $r = -.16$, $p = .10$. Thus, the quality of the individual meta-analyses did not appear to have a substantial influence on estimates of gender differences.

Table 5
Methodological and Statistical Moderators of Gender Differences

Moderator	m	k	d	SD	Range
Quality of methods					
5 (Highest)	34	12,748	.17	.13	.02–.57
4	41	5,983	.22	.12	.02–.51
3	20	1,790	.25	.16	.07–.73
2	6	366	.16	.14	.02–.34
1 (Lowest)	4	200	.28	.19	.16–.56
Sources					
Only published	31	3,709	.24	.14	.02–.51
Some unpublished	65	16,748	.19	.11	.02–.57
Model type					
Fixed-effect	55	10,489	.22	.14	.02–.73
Random-effects	27	4,300	.21	.14	.02–.57
Other	24	6,385	.18	.11	.02–.49
Effect size metric					
d	83	18,237	.21	.13	.02–.73
g	14	2,427	.21	.15	.04–.57
r	9	510	.20	.13	.02–.34

Note. m = number of meta-analyses, k = number of effects, d = absolute effect size, SD = standard deviation of the absolute effect size.

Publication status. Meta-analyses that only incorporate published studies might overestimate the magnitude of gender differences, as studies that obtain null findings might go unpublished. To explore this possibility, we identified whether each meta-analysis incorporated only published ($m = 31$) or both published and unpublished studies ($m = 65$). Meta-analyses that only incorporated published studies yielded somewhat larger effects (.24) than those that incorporated both published and unpublished studies (.19), but this difference was relatively small. In addition, although most of the meta-analyses that we obtained were published articles or chapters ($m = 102$), a few were unpublished dissertations ($m = 4$). Published meta-analyses (effect size $= .21$, $SD = .13$) yielded comparable results to unpublished meta-analyses (effect size $= .16$, $SD = .17$).

Model type. We examined whether estimates of global gender differences fluctuated as a function of the meta-analytic model used to aggregate effects. Along these lines, some meta-analyses used a fixed-effect model ($m = 55$), whereas other meta-analyses used a random-effects model ($m = 27$). Some papers used a different model type (e.g., mixed-effects) or did not provide sufficient information to determine the model used ($m = 24$). Effect size estimates were similar when comparing fixed-effect (.22) to random-effects approaches (.21).

Effect-size metric. Differences between males and females did not fluctuate as a function of the statistical metric used to calculate gender differences. Specifically, meta-analyses using the Cohen's d metric yielded an effect size of .21, meta-analyses using Hedges' g yielded an effect size of .21, and meta-analyses that utilized an r metric yielded a (converted) effect size of .20. The range of effects based on d was much larger than that based on r, but it also involved a ninefold increase in the number of effects generating that range.

General Discussion

We obtained an enormous collection of meta-analytic findings to provide a current, highly comprehensive test of the magnitude of gender differences across domains. The obtained findings were more consistent with the gender similarities hypothesis (Hyde, 2005) than the gender differences hypothesis (Gray, 1992; Tannen, 1991). That is, of the 386 individual meta-analytic effects we obtained, 46.1% were small and 39.4% were very small. Further, when aggregating across the 106 meta-analyses, the average absolute difference between males and females was a small effect (d) of .21, reflecting approximately 84% overlap in distributions of males and females. These findings inform recent debates regarding the overall difference between males and females (Carothers & Reis, 2013; Stewart-Williams & Thomas, 2013).

In addition to estimating the magnitude of gender differences across domains, we examined moderators of gender differences. First, gender differences were somewhat smaller when comparing research on psychological well-being to research on cognitive variables and social/

personality variables, but fewer meta-analyses have been done on well-being. Second, we showed that small gender differences are largely constant across age, culture, and time period. However, prior meta-analyses sometimes did not report the culture of the samples utilized, and more recent meta-analyses typically incorporated both newer and older data. Third, we showed that methodological and statistical factors, such as the methodological quality of prior meta-analyses and type of meta-analytic model used, appear to have only minimal influence on estimates of overall gender differences.

Interpretation of Results

We utilize Cohen's (1988) conventions to interpret the magnitude of gender differences. Lipsey (1990) proposed a similar, but slightly more liberal rubric in which effects of .15 and .45 were labeled as small and medium, respectively. However, it should be noted that effect size cutoff points are somewhat arbitrary and that many if not most effects in psychological science would be considered small in size. For example, foundational research in social psychology on topics including attribution and social influence often yields small effects (Richard et al., 2003), and recent estimates of the effect of violent video games on aggression yielded small effects (Anderson et al., 2010). Therefore, although gender differences are typically small, they should not be regarded as trivial, as even small effects can have important everyday consequences (e.g., passive smoking and lung cancer, calcium intake and bone mass, homework and academic achievement; Bushman & Anderson, 2001). Further, small gender differences may accumulate when summed across domains (Del Giudice, Booth, & Irwing, 2012).

In addition, caution is necessary when interpreting our global effect size estimate of .21. Specifically, our findings should be interpreted as showing generally small differences between males and females on *measures* of psychological outcomes (e.g., tests of math ability), rather than on *actual* psychological outcomes (e.g., actual math ability). It's possible that psychological measures provide imperfect estimates of the constructs they are designed to measure, which could reduce the apparent magnitude of gender differences. Although beyond the scope of the present report, future study should evaluate the degree to which measurement error influences estimates of overall gender differences in metasynthesis (see Schmidt & Oh, 2013).

Finally, although our results provide suggestive evidence that overall gender differences may not fluctuate substantially across age, culture, and generations, we do not believe that these data should be used to infer that gender differences are static or fixed. Indeed, prior research examining gender differences in specific domains has found important fluctuations in gender effects as a function of age, culture, and time period (see Eagly & Wood, 2013; Hyde, 2014). Scholars should continue to examine whether and when these potentially important theoretical variables moderate psychological gender differences both within and across domains.

Limitations and Future Directions

The present findings help establish the overall difference between males and females across domains, but do not pinpoint *why* gender differences occur. Therefore, an important next step for future research will be to evaluate theories regarding why gender differences and similarities occur (e.g., evolutionary, cognitive learning, sociocultural, and expectancy-value theories; see Eagly & Wood, 2013; Hyde, 2014). Additionally, by aggregating data across 106 meta-analyses that span very different content areas, some may find it difficult to interpret our primary model testing gender differences across all domains. To be sure, cross-domain analyses involve more complex interpretations than single-domain analyses. However, it should be noted that the basic question of gender differences across domains has captivated both public and scientific attention for many years (e.g., Gray, 1992; Hyde, 2005). Finally, a limitation of meta-analyses in general is that they sometimes combine studies that are not necessarily comparable due to different populations, methodologies, and conceptualizations of key variables (i.e., the "apples and oranges" problem; Sharpe, 1997). One could argue that this problem is exacerbated in metasynthesis, where researchers combine meta-analyses that address different topics. In the current study, however, this limitation likely did not have an undue influence because gender is typically measured in a parallel fashion across studies.

In demonstrating an overall effect of .21, the current findings suggest that the distributions of males and females on most variables overlap by about 84% (Cohen, 1988). However, the present findings cannot pinpoint the shape of these distributions (i.e., whether or not they are skewed). Furthermore, although our analysis indicates that gender differences are typically small when considering measures of central tendency, gender differences are more pronounced when comparing the tails of male and female distributions. For example, research indicates that gender differences in violence (Daly & Wilson, 1988) and spatial ability (Hedges & Nowell, 1995) are more pronounced when examining the extreme right tails of male and female distributions (i.e., those who score relatively high in violence or spatial abilities). Additionally, research suggests that males are more variable than females on some intellectual and cognitive variables (Machin & Pekkarinen, 2008).

In surveying a large collection of meta-analyses on gender differences, we also uncovered methodological and reporting concerns that should be addressed in future studies. First, researchers should specify the number of males and females they obtained across studies, to allow for better estimates of the total number of participants included in future metasyntheses. Second, researchers should specify the age and culture of the samples obtained and, if possible, examine whether gender differences vary as a function of age and culture. Third, researchers should specify the meta-analytic model they utilized to aggregate findings across studies (e.g., fixed-effect, random-effects), in addition to reporting confidence intervals and standard errors of the observed meta-analytic effects to facilitate formal computations in future metasyntheses. Finally, it is recommended that future meta-analyses examine whether the magnitude of observed gender differences increases after adjusting effects for measurement error in the criterion.

Coda

Hyde (2005) proposed that males and females are far more similar than different. Since then, gender differences and similarities have received an enormous amount of empirical attention. We utilized data from over 20,000 individual studies and over 12 million participants to reevaluate the gender similarities hypothesis and found that its core proposition receives strong support—across most topic areas in psychological science, the difference between males and females is small or very small. However, there are important exceptions where moderate to large gender differences arise. Further, there was a distinct, albeit small difference between males and females when aggregating effects across domains. Thus, although our results suggest that the overall difference between males and females is relatively small, we caution against the conclusion that gender differences are trivial or nonexistent.

In addition, despite substantial progress, future study is still needed before definitive conclusions regarding overall gender differences can be made. Improvements in methodology and reporting in future meta-analyses may provide better estimates of gender differences. Moreover, future research using alternative empirical approaches, such as large population studies (Else-Quest et al., 2010), cross-cultural studies (Schmitt et al., 2012), and archival studies (Daly & Wilson, 1988) should complement our understanding of gender differences. Finally, future research is needed to identify conditions under which gender differences are most pronounced, as well as factors that give rise to gender differences (see Eagly & Wood, 2013; Hyde, 2014). Characterizing gender similarities and differences remains an exciting yet challenging task that should occupy researchers for decades to come.

REFERENCES

American Psychological Association. (2014). *How does the APA define psychology?* Retrieved from http://www.apa.org/support/about/apa/psychology.aspx#answer

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., . . . Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: A meta-analytic review. *Psychological Bulletin, 136,* 151–173. http://dx.doi.org/10.1037/a0018251

Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin, 126,* 651–680. http://dx.doi.org/10.1037/0033-2909.126.5.651

Aydin, A., Sarier, Y., & Uysal, S. (2011). The effect of gender on organizational commitment of teachers: A meta analytic analysis. *Educational Sciences: Theory and Practice, 11,* 628–632.

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin, 137,* 881–909. http://dx.doi.org/10.1037/a0025354

Bonett, D. G. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods, 14,* 225–238. http://dx.doi.org/10.1037/a0016619

Browne, B. A. (1998). Gender stereotypes in advertising on children's television in the 1990s: A cross-national analysis. *Journal of Advertising, 27,* 83–96. http://dx.doi.org/10.1080/00913367.1998.10673544

Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public. Scientific facts versus media misinformation. *American Psychologist, 56,* 477–489. http://dx.doi.org/10.1037/0003-066X.56.6-7.477

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12,* 1–49. http://dx.doi.org/10.1017/S0140525X00023992

Buss, D. M. (2013). The science of human mating strategies: An historical perspective. *Psychological Inquiry, 24,* 171–177. http://dx.doi.org/10.1080/1047840X.2013.819552

Carothers, B. J., & Reis, H. T. (2013). Men and women are from Earth: Examining the latent structure of gender. *Journal of Personality and Social Psychology, 104,* 385–407. http://dx.doi.org/10.1037/a0030437

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

Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist, 67,* 446–462. http://dx.doi.org/10.1037/a0027119

Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin, 137,* 97–130. http://dx.doi.org/10.1037/a0021591

Daly, M., & Wilson, M. (1988). *Homicide.* Hawthorne, NY: Aldine de Gruyter.

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS ONE, 7,* e29265.http://dx.doi.org/10.1371/journal.pone.0029265

Eagly, A. H., & Wood, W. (2013). The nature-nurture debates: 25 years of challenges in the psychology of gender. *Perspectives on Psychological Science, 8,* 340–357. http://dx.doi.org/10.1177/1745691613484767

Else-Quest, N. M., Higgins, A., Allison, C., & Morton, L. C. (2012). Gender differences in self-conscious emotional experience: A meta-analysis. *Psychological Bulletin, 138,* 947–981. http://dx.doi.org/10.1037/a0027930

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103–127. http://dx.doi.org/10.1037/a0018053

Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology, 59,* 981–993. http://dx.doi.org/10.1037/0022-3514.59.5.981

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116,* 429–456. http://dx.doi.org/10.1037/0033-2909.116.3.429

Gorrese, A., & Ruggieri, R. (2012). Peer attachment: A meta-analytic review of gender and age differences and associations with parent attachment. *Journal of Youth and Adolescence, 41,* 650–672. http://dx.doi.org/10.1007/s10964-012-9759-6

Gray, J. (1992). *Men are from Mars, women are from Venus: A practical guide for improving communication and getting what you want in your relationships.* New York, NY: Harper Collins.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269,* 41–45. http://dx.doi.org/10.1126/science.7604277

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. http://dx.doi.org/10.1037/1082-989X.3.4.486

Higgins, J. P. T., Lane, P. W., Anagnostelis, B., Anzures-Cabrera, J., Baker, N. F., Cappelleri, J. C., . . . Whitehead, A. (2013). A tool to assess the quality of a meta-analysis. *Research Synthesis Methods, 4,* 351–366. http://dx.doi.org/10.1002/jrsm.1092

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592. http://dx.doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373–398. http://dx.doi.org/10.1146/annurev-psych-010213-115057

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107,* 139–155. http://dx.doi.org/10.1037/0033-2909.107.2.139

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85,* 616–626. http://dx.doi.org/10.1037/0022-3514.85.4.616

Johnson, B. T., Scott-Sheldon, L. A. J., & Carey, M. P. (2010). Meta-synthesis of health behavior change meta-analyses. *American Journal of Public Health, 100,* 2193–2198. http://dx.doi.org/10.2105/AJPH.2008.155200

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin, 125,* 470–500. http://dx.doi.org/10.1037/0033-2909.125.4.470

Knight, G. P., Guthrie, I. K., Page, M. C., & Fabes, R. A. (2002). Emotional arousal and gender differences in aggression: A meta-analysis. *Aggressive Behavior, 28,* 366–393. http://dx.doi.org/10.1002/ab.80011

Krizan, Z. (2010). Synthesizer 1.0: A varying-coefficient meta-analytic tool. *Behavior Research Methods, 42,* 863–870.

Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review, 11,* 328–363. http://dx.doi.org/10.1177/1088868307302221

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136,* 1123–1135. http://dx.doi.org/10.1037/a0021276

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Machin, S., & Pekkarinen, T. (2008). Assessment. Global sex differences in test score variability. *Science, 322,* 1331–1332. http://dx.doi.org/10.1126/science.1162573

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51,* 93–120. http://dx.doi.org/10.1146/annurev.psych.51.1.93

Maeda, Y., & Yoon, S. (2013). A meta-analysis on gender differences in mental rotation ability measured by the Purdue Spatial Visualization Tests: Visualization of rotations (PSVT:R). *Educational Psychology Review, 25,* 69–94. http://dx.doi.org/10.1007/s10648-012-9215-x

Magnusson, K. (2014, February 3). Interpreting Cohen's *d effect size: An interactive visualization.* Retrieved from http://rpsychologist.com/d3/cohend/

Petersen, J. L., & Hyde, J. S. (2010). A meta-analytic review of research on gender differences in sexuality, 1993–2007. *Psychological Bulletin, 136,* 21–38. http://dx.doi.org/10.1037/a0017504

Reilly, D., & Neumann, D. L. (2013). Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles, 68,* 521–535. http://dx.doi.org/10.1007/s11199-013-0269-0

Richard, F. D., Bond, C., & Stokes-Zoota, J. J. (2003). One hundred years of Social Psychology quantitatively described. *Review of General Psychology, 7,* 331–363. http://dx.doi.org/10.1037/1089-2680.7.4.331

Riley, J. L., III, Robinson, M. E., Wise, E. A., Myers, C. D., & Fillingim, R. B. (1998). Sex differences in the perception of noxious experimental stimuli: A meta-analysis. *Pain, 74,* 181–187. http://dx.doi.org/10.1016/S0304-3959(97)00199-1

Schmidt, F. L., & Oh, I. (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes, 121,* 204–218. http://dx.doi.org/10.1016/j.obhdp.2013.03.002

Schmitt, D. P., Jonason, P. K., Byerley, G. J., Flores, S. D., Illbeck, B. E., O'Leary, K. N., & Qudrat, A. (2012). A reexamination of sex differences in sexuality: New studies reveal old truths. *Current Directions in Psychological Science, 21,* 135–139. http://dx.doi.org/10.1177/0963721412436808

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17,* 881–901. http://dx.doi.org/10.1016/S0272-7358(97)00056-1

Shuster, J. J. (2010). Empirical vs natural weighting in random effects meta-analysis. *Statistics in Medicine, 29,* 1259–1265. http://dx.doi.org/10.1002/sim.3607

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis

allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

Stewart-Williams, S., & Thomas, A. G. (2013). The ape that thought it was a peacock: Does evolutionary psychology exaggerate human sex differences? *Psychological Inquiry, 24,* 137–168. http://dx.doi.org/10.1080/1047840X.2013.804899

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135,* 859–884. http://dx.doi.org/10.1037/a0017364

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

Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance a meta-analysis. *Psychological Bulletin, 98,* 260–282. http://dx.doi.org/10.1037/0033-2909.98.2.260

Twenge, J. M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles, 36,* 305–325. http://dx.doi.org/10.1007/BF02766650

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. http://dx.doi.org/10.1037/0021-843X.111.4.578

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9,* 111–125. http://dx.doi.org/10.1177/1745691613518075