**Modeling and Explaining Sex Differences in the Prevalence of Autism Spectrum Disorder:
A Meta-Analytic Approach**

Meng-Ting Chen

Department of Child and Family Studies, Fu Jen Catholic University
510 Zhongzheng Rd, Xinzhuang Dist, New Taipei City 24205, Taiwan
montine729@gmail.com, +886-939-605-325

Yen-Ping Chang

Institute of Sociology, Academia Sinica
HSSB South Rm 1034, 128 Academia Rd Sec 2, Nankang Dist, Taipei City 11529, Taiwan
yenpingchang@hotmail.com, +886-988-025-065

Xiaopeng Lu

Language Technologies Institute, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213, USA
xiaopen2@andrew.cmu.edu, +1-919-903-7850

Rune J. Simeonsson

School of Education, University of North Carolina at Chapel Hill
UNC CB 3500 Peabody Hall, Chapel Hill, NC 27599, USA
rjsimeon@email.unc.edu, +1-919-962-2512

Marisa E. Marraccini

School of Education, University of North Carolina at Chapel Hill

UNC CB 3500 Peabody Hall, Chapel Hill, NC 27599, USA

mmarracc@email.unc.edu, +1-919-843-5099

Correspondence

Yen-Ping Chang, Institute of Sociology, Academia Sinica, HSSB South Rm 1034, 128
Academia Rd Sec 2, Nankang Dist, Taipei City 11529, Taiwan; yenpingchang@hotmail.com;
+886-988-025-065

Authors' copy of manuscript under review for publication since 3/10/2020

Abstract

Why is autism spectrum disorder (ASD) less prevalent among females than among males? Synthesizing existing theories, we constructed two statistical models, one for each of both classes of theories on the issue, and derived testable competing predictions for the expression hypothesis—females express less severe ASD traits so are diagnosed less—against the perception hypothesis—females’ expressions are perceived as less severe so are diagnosed less—in their associations with expected symptom severity of ASD-diagnosed individuals. Examining the predictions with the highest precision, a meta-analysis on 103,958 historical assessment scores across ages, IQ, diagnoses, and assessments indicated that diagnosed females and males show almost equally severe and non-conditioned symptoms. Given the models are direct representations of the literature, the evidence therefore lends strong support for both, but not neither of the hypotheses, showing that the diagnosis of ASD is simultaneously an expression and a perception issue of the disorder.

Keywords: autism spectrum disorder, sex, gender, meta-analysis

Autism spectrum disorder (ASD)—a neuro-developmental disorder characterized by the three symptoms of communication difficulty, social interaction difficulty, as well as repetitive and restrictive behaviors and interests—is known for being much more prevalent among males than among females. Although studies do not agree on the exact degree of this difference between sexes (1, 2), the overall existence of such disparity in the prevalence of ASD has been shown across age groups (3, 4), ranges of intelligence quotient (IQ), diagnostic criteria, geographical regions (5), and by multiple independent research teams (6).

Despite its well-documented phenomenology, the field remains fragmented with regard to what causes this sex difference in ASD diagnostic prevalence (3, 4, 7), and two general classes of explanations have been proposed in the literature (8). Although not logically contradictory, the two approaches are largely opposite in spirit, with one focusing on the bio-neural essence of ASD while the other originating from potential social construction of the disorder. As such, the two lines of reasoning have long been tested individually and never empirically integrated. Contributing to an integration, we tackle the causes of the sex difference in ASD diagnostic prevalence in the present research.

The Expression Hypothesis

Aware of the diversity and specificities of the theories, we label the first class of theories concerning individuals' expressions of ASD symptoms the *expression hypothesis*, modeling it as follows ¹. Some researchers have theorized that, one way or another, females appear inherently less susceptible to ASD than are males. Females express ASD symptoms with less severity, and hence are diagnosed with the disorder less frequently. For example, the Imprinted-X Liability Threshold Model suggests that females may develop ASD less often because only females, but not males, inherit the paternal X-chromosome and its capability to suppress ASD-symptom-like behaviors (9). Similarly, the Extreme Male Brain Theory argues that the female brain is equipped with stronger “empathizing”—the ability to see the world from others’ perspectives—while the male brain is wired with stronger “systemizing”—the ability to find patterns in the environment. At the extreme, overly male brains might then be constrained from thinking empathically, causing communication and social interaction difficulty, and from breaking the rigid systems of behavioral rules that the male brain has constructed for itself to understand and interact with the world, causing repetitive and restrictive behaviors and interests (10).

¹ Given that our theorizing is more statistically driven (as shown below) and, therefore, might not map squarely on that of 8.L. Hull, W. Mandy, Protective effect or missed diagnosis? Females with autism spectrum disorder. *Future Neurology* **12**, 159-169 (2017)., we did not adopt these researchers’ terminology and instead focused more on naming classes of theories by their implications for their statistical models used in the present work.

We suggest that, because females appear to suffer fewer “points of weakness” for developing ASD than do males following theories under the expression hypothesis (9), severity of females’ autistic traits in the general population can be modeled using a chi-square distribution ² with fewer degrees of freedom (*df*; the red curve F in Fig. 1) to deviate from the norm. By comparison, males’ autistic trait severity can be modeled as a chi-square distribution with more *dfs* (the blue curve M). Consequently, because the smaller the *df* of a chi-square distribution, the smaller the area under the curve in the (right) tail of the distribution (the colored areas); females (the red area) should be less likely to be found in the smaller tail of their distribution than should males in their larger tail (the red and the blue area combined). In other words, females are expected to be less likely to exceed the severity threshold (T of F/M) that defines the tails of distributions and serves as the diagnostic criterion of ASD for both sexes. As a result, the current model of the expression hypothesis predicts that females receive the diagnosis of ASD less often than do males, in line with the existing research.

² There are several reasons why the chi-square distribution, as opposed to other types of distributions, is a good starting point for modeling the expression hypothesis. Theoretically, the chi-square distribution is invented to describe the probability distribution of the level of deviation in the population from the norm, as in the case of autistic traits understood as a deviation from the typical. Specifically, this probability of deviation is derived from the assumption that the issue at hand is governed by a finite number—the *df*—of underlying risk factors that independently contribute to the total deviation of a sample, an assumption fits squarely with the field’s current understanding that ASD likely has multiple unique causes (as reviewed). Phenomenologically, the chi-square distribution is a single-modal distribution with a positive (right) tail. This may then reflect the fact that most individuals in the population are typical—with a less severe autistic traits—while others show more severe autistic traits, sitting on the tail of the distribution and potentially diagnosed with ASD thereof. In addition, the literature has indicated that the non-clinical, typical, population may share the same etiology of ASD with those who have been diagnosed with ASD 11.J. N. Constantino, R. D. Todd, Autistic Traits in the General Population: A Twin Study. *Archives of General Psychiatry* **60**, 524-530 (2003). This thus precludes the use of multi-modal distributions to represent autistic traits (e.g., by giving typical individuals a larger mode on the low end and autistic individuals another smaller mode on the high end) and favors single-modal distributions such as the chi-square distribution. Finally, empirical evidence supports the use of the chi-square distribution too, when revealing that autistic traits do seem to approximate to the chi-square distribution in the population, more so than to other commonly used distributions 11. J. N. Constantino, R. D. Todd, Autistic Traits in the General Population: A Twin Study. *Archives of General Psychiatry* **60**, 524-530 (2003); 12.

A. Abu-Akel, C. Allison, S. Baron-Cohen, D. Heinke, The distribution of autistic traits across the autism spectrum: evidence for discontinuous dimensional subpopulations underlying the autism continuum. *Molecular Autism* **10**, 24 (2019)..

Figure 1. An illustration of the distribution of autistic-trait severity in the population following the expression hypothesis. F, M, and T(F/M) designate the distributions of females, of males, and the diagnostic threshold for both sexes. The units of both axes are arbitrary and only for illustration purposes.

Figure 2. An illustration of the distribution of autistic-trait severity in the population following the perception hypothesis. F/M, T(F), and T(F) designate the distribution of both sexes, the diagnostic thresholds for females, and for males. The units of both axes are arbitrary and only for illustration purposes.

The Perception Hypothesis

The other class of theories on sex disparity in ASD prevalence focuses on how the expression of symptoms is socially perceived; we label this class of theories *the perception hypothesis*. According to the theories, one way or another, females are less likely to be recognized as atypically autistic than are males, even if both sexes manifest an equal level of symptom severity. Females are therefore less susceptible to the *labeling* of ASD and diagnosed with the disorder less often. For instance, Lai, Lombardo, Auyeung, Chakrabarti and Baron-Cohen (3) reviewed the literature and suggested that females with ASD might be more frequently missed or misdiagnosed—i.e., labeled with disorders other than ASD—than are their male counterparts, because society defines normality differently for the sexes and judges them against incompatible standards. For example, women may be expected to be submissiveness, not assertive or interested in changing others' opinions in social interactions and conversations, and thus to be regarded as more or less naturally ASD-like compared to men. Women's autistic traits might then be more accepted and even encouraged. Echoing this observation, Dworzynski, Ronald, Bolton and Happé (13) reported that, without intellectual disabilities or additional behavioral issues, females are less likely to receive the diagnosis of ASD than are males, even if both parties present similarly severe autistic characteristics. To put it another way, relatively, females with ASD need to evidence intellectual or other challenges to be considered for an ASD diagnosis, while their male counterparts do not. Finally, previous research has revealed that, even when females' ASD is formally recognized and diagnosed, it is often recognized later in life than it is for males (14). Overall, females may either have to fulfill extra criteria, or demonstrate ASD symptoms at a higher frequency and/or intensity to receive the same label of ASD as do males (15).

In terms of modeling the perception hypothesis—different from modeling the expression hypothesis, which operates on the *df* of the distribution of autistic-trait severity—we allow the diagnostic threshold to vary between sexes while holding the *dfs* of their distributions a constant. That is, both sexes are assumed to be inherently equally affected by ASD in the population, following the same severity distribution (the black curve F/M in Fig. 2). Nevertheless, there is a higher threshold (T of F) and, thus, a smaller tail of diagnosis for females (the red area) than for males (T of M, and the red and the blue area combined, respectively). Consequently, females are again expected to be less likely to exceed “their” diagnostic criterion of ASD than are males against their other, lower criterion. As a result, the current model of the perception hypothesis similarly predicts that females receive the diagnosis of ASD less often than do males, as was predicted by the model of the expression hypothesis and documented in the literature.

Competing Predictions of the Two Hypotheses

Given that both models above and the theories they represent predict similar sex disparity in ASD prevalence from rather distinct understandings of the nature of the diagnosis, the question that remains is what distributional characteristics of the models would differentiate the expression hypothesis from the perception one and, at the same time, be practically accessible. For the literature, the issue of practicality may be of special importance because one major reason that the two classes of hypotheses were commonly separate is that it is extremely difficult to empirically examine the perception hypothesis, not to say comparing it with the expression one, as the theories under the perception hypothesis often concern possible underrepresentation of females, or, in other words, female participants who are not in researchers' datasets and, therefore, cannot be studied.

Contributing to solving the predicament, we saw an opportunity to rely on unsampled females' indirect traces in the sampled data to make meaningful inferences about them, using the statistical models proposed above. Specifically, we examine the severity of ASD symptoms of those who have been diagnosed—those who have been sampled. For the models, this means an analysis of the average severity (score) of only those who exceed their diagnostic thresholds and have entered the tails of their distributions. Particularly, the expression hypothesis predicts females' average severity in the tail (the red area in Fig. 1) to be lower than that of males (the red and the blue area combined). It is because, relative to males, diagnosed females tend to have “just passed” the diagnostic threshold (T of F/M) and crowd on the low (left) side of the tail. On the other hand, given the heavier tail of males' distribution than that of females', diagnosed males tend to be farther away from the threshold and, thus, demonstrate more severe symptoms than do their female counterparts. Consequently, the model of the expression hypothesis suggests that females diagnosed with ASD show lower average symptom severity than do diagnosed males.

Opposing that of the expression hypothesis, the model of the perception hypothesis nonetheless predicts females' average severity in the tail (the red area in Fig. 2) to be higher than that of males (the red and the blue area combined). It is because the higher diagnostic threshold for females (T of F) than that for males (T of M) prevents females whose severity is in between both thresholds from receiving a diagnosis and from counting toward the average severity of diagnosed females. The same intermediate-severity group of males, however, can receive the diagnosis of ASD and lower the average severity of diagnosed males, because of males' lower diagnostic threshold. Consequently, the model of the perception hypothesis suggests that females diagnosed with ASD show higher average symptom severity than do diagnosed males—opposite to what the expression hypothesis would suggest. Inspecting the average difference in symptom severity between diagnosed females and males may then help differentiate the expression

hypothesis from the perception hypothesis and shed light on why ASD is more prevalent among males than among females.

The Current Investigation

To date, it remains unclear whether the sex disparity in ASD prevalence is driven by sexually differential expression of ASD symptoms, or by sexually differential perception given to such expression. Aiming at comparing the hypotheses, the theoretical models we presented derive competing predictions of the expression hypothesis and the perception hypothesis, focusing on whether females diagnosed with ASD show less (following the expression hypothesis) or more (following the perception hypothesis) severe ASD symptoms than do males diagnosed with ASD. Following the models, the purpose of the current study was to conduct a comprehensive empirical test—a meta-analysis³—that examined differences in the severity of ASD symptoms between diagnosed females and males. If, in aggregate, diagnosed females were reported to show less severe ASD symptoms than were their male counterparts, the evidence would favor the expression hypothesis over the perception one. By contrast, if diagnosed females were reported to display more severe symptoms than were their male counterparts, the evidence would favor the perception hypothesis over the expression one.

It is worth noting here, however, the two hypotheses that we focus on are not theoretically exclusive. Females and males might be affected by ASD differently (the expression hypothesis), while simultaneously being assigned different diagnostic criteria (the perception hypothesis). As such, the goal of the present research was to estimate the size of the effect associated with each hypothesis, gauging one against the other and, thus, empirically linking them together.

Further, we considered the possibility that difference in ASD severity between diagnosed females and males is conditioned to and varies between specific aspects of ASD and of individuals. For instance, ASD has three rather unique core symptoms (*1*), making it possible that the degree of the target sex disparity changes between symptoms. Further, such a distinction might only exist in some but not all ranges of age (*4*) or across varying cognitive abilities as measured by IQ (*16, 17*). Lastly, we explored the moderation effect of the diagnostic systems (e.g., Diagnostic and Statistical Manual of Mental Disorders [DSM] or International Classification of Diseases [ICD]) that were used to

³ Even if both the expression and the perception model are not normally distributed—let alone their truncated tails we suggest analyzing here—the estimation of the average sex difference in the tails, that is, its sampling distribution, should still approximate to a normal distribution according to the central limit theorem. This justifies our empirical use of meta-analysis to test the models.

guide individuals' diagnoses in the first place, and a similar effect of the assessment tools (e.g., Autism Diagnostic Observation Schedule [ADOS], Autism Diagnostic Interview [ADI], or others) that were applied to measure individuals' symptom severity scores. In summary, the present research meta-analyzed the main effect of sex on ASD symptom severity of diagnosed individuals, as well as the moderation effects of symptom type, age, IQ, diagnostic criterion, and assessment tool on this main effect of sex. Below, we detail why, and then how we conducted a meta-analysis to address these effects.

Method

We employed a meta-analysis to compare ASD severity between diagnosed females and males, because the meta-analytical approach is stronger in many ways than is the conventional one, that is, recruiting a new independent sample of subjects to test a research question. For instance, by pooling weighted estimates of effects, meta-analysis achieves much greater statistical power than does any of its included individual empirical studies. This is especially helpful for addressing a research question focused on a unique subgroup within the population (e.g. individuals with ASD) and hence regularly faces limited power. In addition, small sample sizes are not merely a statistical problem, but also a theoretical one, as they increase the risk of biased sampling and, subsequently, biased inferences. By contrast, meta-analysis is an efficient and promising method to answer a range of existing as well as emerging questions. For that reason, Van Wijngaarden-Cremers, van Eeten, Groen, Van Deurzen, Oosterling and Van der Gaag (4) conducted a meta-analysis similar to ours on the differences in the severity of symptoms between females and males diagnosed with ASD. The findings of the research are more robust and comprehensive than are previous studies (18). This preliminary meta-analysis, however, was limited in statistical power because the authors separated analyses into three individual models addressing the three core symptoms of ASD. Furthermore, the researchers did not account for several potentially meaningful moderators (e.g., those reviewed above), missing the opportunity to parse out the heterogeneity they themselves discovered among the studies. As a result, it is still an open question as to whether and how sex differences in ASD symptom severity would change once age, IQ, diagnostic criteria, assessment tools used to measure symptom severity, and the specificities of symptoms have been taken into account. Lastly and critically, Van Wijngaarden-Cremers, van Eeten, Groen, Van Deurzen, Oosterling and Van der Gaag (4) did not associate their findings with theoretical explanations for the findings, instead presenting their work as a quantitative summary of the literature. That is, the authors merely sought to answer whether ASD-diagnosed females and males differ in their symptom severity, overlooking that the answer to the question may further explain whether it is the expression or the

perception of ASD symptoms that causes the sex difference in the prevalence of ASD. As such, the current investigation not only extended the literature, the empirical endeavor of Van Wijngaarden-Cremers, van Eeten, Groen, Van Deurzen, Oosterling and Van der Gaag (4), but indeed tackled a distinct theoretical question.

Literature Search

We conducted a systematic meta-analysis followed the PRISMA guidelines on studies that examined the differences between females and males diagnosed with ASD, in their symptom severity. Two exclusive literature searches were carried out to form the dataset for analysis. First, we included all studies analyzed in the meta-analysis of Van Wijngaarden-Cremers, van Eeten, Groen, Van Deurzen, Oosterling and Van der Gaag (4), assuming that the research had identified the studies needed for our current investigation and published before June 2013 (i.e., the scope of Van Wijngaarden-Cremers and colleagues' meta-analysis). We then supplemented this portion of the data (including results up to June, 2013) by conducting our own search for empirical findings published after June 2013 and before December 2017 in the following bibliographic databases: PsycINFO, Scopus, PubMed, and Web of Science. To be overly inclusive at this stage, the following sets of keywords were used: 1) autism spectrum disorder, ASD, autism, pervasive development disorder, or PDD; and 2) sex, gender, or female(s); and 3) IQ, intelligence, cognitive, or cognition. Articles were evaluated based on the following inclusion criteria: 1) the article reported at least one sex comparison in the severity of any of the three core symptoms of ASD based on ADOS or ADI-R assessment tools or of their composite; and 2) the article provided enough information for extracting the effect size of the comparison.

As illustrated in Figure 3, our search initially identified 2,806 articles published after June 2013. Their titles and abstracts were screened by the lead author and four trained undergraduate research assistants for relevance; the screening resulted in 84 articles for full-text review. Of these, 28 met the inclusion criteria, thus leaving a total of 49 articles—28 from our search and 21 from the analysis of Van Wijngaarden-Cremers, van Eeten, Groen, Van Deurzen, Oosterling and Van der Gaag (4)—for the present research. A second and final review of all these articles was conducted for interpretability, resulting in the exclusion of five articles because their participants either were not formally diagnosed (N of articles = 2), had comorbid genetic disorders (N = 1), gender dysphoria (N = 1), or the researchers used an autistic-traits-matched sample so all reported sex differences were already artificially balanced out (N = 1). As a result, the final dataset for analysis was composed of 43 articles—84,718 data points of ASD

symptoms from on average 9,192.00 males and 19,240 points from on average 2,412.17 females ⁴—among which 39, 33, and 32 reported scores in the social, communication, and the repetitive behavior domain, respectively.

⁴ These numbers of individuals were not integers because sometimes there were multiple measures in a study and participants might miss some but not all those measures. Effectively, then, the individuals did not each complete one full set of data.

Figure 3. The PRISMA flow diagram of the present meta-analysis.

Data Preparation

Given that sex is commonly reported as a binary variable in the literature, Cohen's d was derived for each comparison between sexes in any of the three core symptoms or their composite. This effect-size extraction was repeated for all comparisons in the participant sample reported in an article, if there were multiple samples in an article or multiple measures for a sample. Here, it is worth noting that even though the data were in a conceptually three-level structure, wherein effect sizes of comparisons were nested in samples that were themselves nested in articles, we ignored the level of articles because only two studies recruited more than one (i.e., two) independent samples of participants.

Subsequently, we coded each effect size based on the types of symptoms that it covered—it might cover more than one—and the assessment instrument used to generate the effect, as well as each participant sample's age, IQ, and diagnostic criteria. Specifically, we coded each effects size as either 0 or 1 in each of the three core symptoms (i.e., communication impairment, social impairment, and repetitive/restrictive behaviors), with 1 identifying effect sizes that did *not* measure the symptom and 0 identifying effect sizes that did measure the symptom. For example, for the communication subscale of Vineland, the scale was coded 0 in its target domain of communication impairment, and 1 in both social impairment and repetitive/restrictive behaviors. The codes were developed this way, contrary to the convention of using 1 for construct presence, because our three codes would together make the intercept of the model—where all three symptoms were 0 and thus “present” in our case—the size of the theoretical composite ASD severity. By contrast, if we had followed the tradition and reversed the codes, the intercept would estimate the uninterpretable effect of “no symptom measured.”

As for moderator coding, a participant sample was grouped by its mean age into a subsample of either infants and toddlers (sample mean age 0-3 years old; $N = 3$), preschoolers (3-6 years old; $N = 4$), children (6-12 years old; $N = 19$), adolescents (12-18 years old; $N = 13$), or adults (older than 18 years old; $N = 6$), with the children group being the largest and thus the reference group. Regarding IQ, participants were categorized as having either a mean IQ less than 70 (the reference group), or greater than 70 (without intellectual disability). For studies that only provided scores of mental ages without an IQ, the mental-age-equivalent IQ was derived by dividing the mean mental age by chronological age. Regarding symptom assessments, due to the large number of assessments used by the included studies (i.e., 25), we combined all versions of ADOS into one group (the reference group), and all versions of ADI into another. The two groups weighted about equally and highly, together accounting for 84% of the effect sizes. We

therefore merged all other assessments into a third, “others,” group. Lastly, diagnostic criteria were coded by separating the DSM (the reference group) from the ICD system, ignoring versions of the systems, because only a few studies used versions of the diagnoses other than the popular DSM-4, DSM-5, and ICD-10. It is important to note that some studies did not indicate their diagnostic criteria at all, or ambiguously stated that they applied some mixture of DSM and ICD. We excluded these studies when analyzing diagnostic criteria, leaving the analysis a smaller dataset (K of effect sizes = 199, cf $K = 293$ for others).

Analytic Plan

Using the *metafor* package of R (19), in four steps, we analyzed the data in a two-level random-effect meta-regression wherein effect sizes were nested within participant samples. To begin, funnel plots and the trim-and-fill method were used to assess and, where necessary, address potential outliers and publication bias (see Outliers in Results). By calculating the Q -statistic, homogeneity of variance was tested among effect sizes; we expected it to be significant based on previous research (4). After inspecting the data, in the second step, we estimated the main effect of sex on the overall severity of ASD, that is, the intercept of the intercept-only meta-regression. Building upon this baseline model, in the third step, the distinctive effects of the three symptoms were then modeled by adding the three symptom-type codes in the meta-regression. Finally, non-symptom moderators (i.e., including age, IQ, diagnostic criteria, and assessment) were entered one at a time into this symptom-specific model as “Moderator x Symptom” interactions. Simple effects were derived for significant moderations found.

Results

Steps 1 and 2: Data Inspection and Non-Moderated Overall Results

Outliers. We used funnel plots to identify outliers among the 295 effect sizes ($K = 295$) collected from 45 participant samples. The plots (Figure 4A) indicated that there were two possible outliers among the effects: the one with the lowest effect size and the highest standard error (the bottom left), and the other with the highest effect size and the second highest standard error (the bottom right). To be cautious, we first fit a two-level random-effect model without moderators to all the effects including these potential outliers before ($K = 295$); we found that the effects were significantly heterogeneous ($Q = 1757.020, p < .001$), with no significant overall gender difference in composite ASD severity across all effects not otherwise distinguished ($d = 0.047, Z = 0.423, p < .672$) (see Table 1). The same model was then fit to the cleaned data without the two outliers (Figure 4B; $K = 293$); again (Table 1), significant heterogeneity

was found ($Q = 1749.719$, $p < .001$) without significant overall sex difference in ASD severity ($d = 0.013$, $Z = 0.194$, $p < .001$).

Trim-and-fill process. After removing outliers, visual inspection of the funnel plot (Figure 4B) suggested a publication bias toward reporting positive effects—the funnel missed the bottom left corner—even though such a bias did not result in a significant sex difference. To be comprehensive, the trim-and-fill method was still applied to the cleaned data. The results (Table 1 and Figure 4C) consolidated the finding that there was no significant sex difference in composite ASD severity ($d = -0.046$, $Z = -0.654$, $p < .001$) when potential missing effects were simulated and added back ($K = 313$).

Table 1.

Results of Meta-Analyses of ASD Sex Deference

Model	Est.	SE	Z	p of Z	[CI 95% bounds]		Q	p of Q
K = 295	0.047	0.112	0.423	.672	-0.172	0.267	1757.020*	.001
K = 293	0.013	0.068	0.194	.846	-0.121	0.148	1749.719*	.001
K = 313	-0.046	0.071	-0.654	.513	-0.186	0.093	1830.298*	.001

Note: Model K = 295 is the original dataset; model K = 293 is without two outlier effects; model K = 313 is trimmed and filled from model K = 293.

Figure 4. Funnel plots. Plot A is of all effects collected; Plot B is of all effects except the two outliers marked in Plot A; Plot C is Plot B after the trim-and-fill.

Steps 3: Symptom-Specific Effects

The results from the random-effects model (Table 1) did not support significant overall differences between females and males with an ASD diagnosis in composite ASD severity; however, significant heterogeneity of effect sizes was supported by the Q statistic. Given no sex difference was found, the significant heterogeneity (Q statistic), and that moderation analysis cannot be conducted on simulated trimmed-and-filled data (because they do not have values on the moderators), planned moderation analyses were conducted on the pre-trim-and-fill data without the identified outlier (data $K = 293$).

All the effects except the outliers were fit to the same two-level random-effects meta-regression where effects were nested within participant samples and moderated by the three symptom codes. We found that (Table 2) sex differences significantly varied between ASD core symptoms in that ASD severity was significantly greater for males compared to females when social or communication impairments were not considered in a measure. Upon closer examination, the simple effects analysis revealed that, when only assessing one symptom at a time (using all data including those of non-focal symptoms; the same all-data approach was used for all focused analyses below), there were no significant sex differences in social (est. = -0.100 ; SE = 0.140 , $Z = -0.718$, $p = .473$, CI_{95%} = $[-0.374, 0.174]$) and communicative impairment (est. = -0.044 ; SE = 0.144 , $Z = -0.305$, $p = .760$, CI_{95%} = $[-0.327, 0.239]$). Yet males with ASD showed greater restricted and repetitive behavior than did their female counterparts (est. = 0.462 ; SE = 0.147 , $Z = 3.147$, $p = .002$, CI_{95%} = $[0.174, 0.750]$).

Table 2.

Results of Meta-Analysis of ASD Symptoms

Predictor	Est.	SE	Z	p	[CI _{95%} bounds]	
Residual Q *	1514.968			<.001		
Intercept	-0.312	0.309	-1.009	.313	-0.917	0.293
S *	0.415	0.177	2.351	.019	0.069	0.761
C *	0.359	0.175	2.045	.041	0.015	0.703
B	-0.147	0.194	-0.759	.448	-0.528	0.233
Simple effect						
S	-0.100	0.140	-0.718	.473	-0.374	0.174

C	−0.044	0.144	−0.305	.760	−0.327	0.239
B *	0.462	0.147	3.147	.002	0.174	0.750

Note: S = social impairment; C = communicative impairment, B = restricted and repetitive behavior; * = $p < .05$.

Step 4: Moderators

IQ. As shown in Table 3, results indicated that there was no significant moderation effect of intellectual disability (ID) on the overall sex difference in ASD severity. The ID x Symptom interaction was not significant either. The model replicated the found general pattern (Table 2) that ASD severity was higher for males than for females only in repetitive behavior, but not other two symptoms or the composite severity.

Table 3.

Results of Meta-Analysis of Intellectual Disability (ID; v. IQ ≥ 70 ; K=288)

Predictor	Est.	SE	Z	p	[CI 95% bounds]	
Residual Q *	1291.894			<.001		
Intercept (IQ ≥ 70)	−0.353	0.344	−1.026	.305	−1.027	0.321
S *	0.426	0.195	2.190	.029	0.045	0.808
C *	0.426	0.193	2.207	.027	0.048	0.804
B	−0.088	0.214	−0.414	.679	−0.508	0.331
ID	0.368	0.820	0.449	.654	−1.240	1.975
x S	0.022	0.491	0.045	.964	−0.940	0.984
x C	−0.507	0.494	−1.025	.305	−1.476	0.462
x B	−0.609	0.551	−1.105	.269	−1.688	0.471

Simple effect

Baseline (IQ ≥ 70)

S	−0.016	0.155	−0.102	.919	−0.319	0.288
C	−0.015	0.166	−0.091	.928	−0.340	0.310
B *	0.499	0.165	3.024	.002	0.176	0.823

Note: S = social impairment; C = communicative impairment, B = restricted and repetitive behavior; * = $p < .05$.

Age. As shown in Table 4, the results indicated that age only affected sex differences in ASD when restricted and repetitive behavior was removed from the composite ASD severity for samples that were equal to or older than 18. To probe this interaction, simple effects for this age group were derived, revealing that the composite ASD severity was significantly higher for female than for male adults (est. = -2.382 ; SE = 0.759 , $Z = -3.138$, $p = .002$, CI_{95%} = $[-3.870, -0.884]$). Nonetheless, the same sex difference did not appear for individual symptoms (social impairment est. = 0.423 ; SE = 0.412 , $Z = 1.026$, $p = .305$, CI_{95%} = $[-0.385, 1.231]$; communicative impairment est. = 0.534 ; SE = 0.436 , $Z = 1.222$, $p = .222$, CI_{95%} = $[-0.322, 1.389]$; restricted and repetitive behavior est. = 0.424 ; SE = 0.380 , $Z = 1.115$, $p = .265$, CI_{95%} = $[-0.321, 1.170]$).

Table 4.

Results of Meta-Analysis of Age (v. age 6-12; K=293)

Predictor	Est.	SE	Z	p	[CI _{95%} bounds]	
Residual Q *	1063.849			<.001		
Intercept (age 6-12)	0.137	0.408	0.335	0.738	-0.663	0.936
S	0.085	0.259	0.328	0.743	-0.423	0.593
C	-0.513	0.281	-1.828	0.068	-1.064	0.037
B	0.277	0.254	1.089	0.276	-0.221	0.774
Age 0-3	1.708	2.213	0.772	0.440	-2.630	6.046
x S	-0.620	1.073	-0.578	0.563	-2.722	1.482
x C	-1.171	1.253	-0.934	0.350	-3.627	1.286
x B	-1.305	1.165	-1.120	0.263	-3.588	0.978
Age 3-6	-0.353	1.415	-0.249	0.803	-3.126	2.42
x S	0.436	0.883	0.494	0.621	-1.294	2.167
x C	0.032	1.138	0.028	0.977	-2.198	2.263
x B	-0.289	0.883	-0.327	0.743	-2.019	1.441
Age 12-18	-0.058	0.962	-0.061	0.952	-1.944	1.827
x S	0.266	0.530	0.502	0.616	-0.773	1.305
x C	0.259	0.549	0.472	0.637	-0.817	1.335
x B	-0.105	0.526	-0.200	0.842	-1.136	0.926
Age 18+ *	-2.519	0.862	-2.923	0.003	-4.208	-0.830

x S *	1.373	0.538	2.553	0.011	0.319	2.428
x C *	1.971	0.578	3.409	<.001	0.838	3.104
x B *	1.071	0.529	2.025	0.043	0.035	2.108
<hr/>						
Simple effect						
Baseline (age 6-12)						
S	-0.100	0.239	-0.419	0.675	-0.569	0.369
C	-0.292	0.253	-1.153	0.249	-0.788	0.204
B *	0.498	0.236	2.110	0.035	0.035	0.961
Age 18+ *	-2.382	0.759	-3.138	.002	-3.870	-0.884
S	0.423	0.412	1.026	.305	-0.385	1.231
C	0.534	0.436	1.222	.222	-0.322	1.389
B	0.424	0.380	1.115	.265	-0.321	1.170

Note: S = social impairment; C = communicative impairment, B = restricted and repetitive behavior; * = $p < .05$.

Diagnostic criteria. As shown in Table 5, the results indicated that only the DSM diagnostic system replicated the general pattern of sex differences in ASD symptoms: Males with ASD showed greater restricted and repetitive behavior than did females with ASD. No other difference was found, and the ICD showed no sex difference at all, in either of the three symptoms and their composite.

Table 5.

Results of Meta-Analysis of Diagnosis (v. DSM; K = 199)

Predictor	Est.	SE	Z	p	[CI _{95%} bounds]	
Residual Q *	783.200			<.001		
Intercept (DSM)	0.070	0.250	0.282	.778	-0.420	0.561
S *	0.276	0.140	1.976	.048	0.002	0.550
C	0.146	0.137	1.062	.288	-0.123	0.415
B	-0.276	0.155	-1.780	.075	-0.579	0.028
ICD *	-2.712	1.349	-2.010	.044	-5.355	-0.068
x S *	1.638	0.713	2.298	.022	0.241	3.035
x C	1.167	0.714	1.635	.102	-0.232	2.566

x B	1.141	0.714	1.599	.110	−0.257	2.540
Simple effect						
Baseline (DSM)						
S	−0.060	0.143	−0.416	.678	−0.340	0.221
C	0.071	0.150	0.471	.638	−0.224	0.366
B *	0.492	0.149	3.308	<.001	0.201	0.784
ICD *	−2.641	1.325	−1.993	.046	−5.239	−0.043
S	−0.463	0.290	−1.599	.110	−1.031	0.104
C	0.139	0.320	0.433	.665	−0.489	0.766
B	0.586	0.324	1.806	.071	−0.050	1.221

Note: S = social difficulty; C = communicative difficulty, B = restricted and repetitive behavior; * = significant.

Assessment tool. The results demonstrated that (Table 6) ADOS, the reference assessment, replicated the overall pattern of ASD symptom severity, showing only such a disparity in restricted and repetitive behavior, but not other individual symptoms or the composite. We further discovered that there was no significant deviation of ADI from ADOS and that assessment tools other than ADOS and ADI (i.e., the others group) together differed from ADOS. Specifically, although males assessed by these other tools also showed significantly greater restricted and repetitive behavior than did females, they now also showed significantly lesser social impairment, communication impairment, and composite severity.

Table 6.

Results of Meta-Analysis of Assessment Tool (v. ADOS; K = 293)

Predictor	Est.	SE	Z	p	[CI _{95%} bounds]	
Residual Q *	1413.319			<.001		
Intercept (ADOS)	0.539	0.452	1.192	.233	−0.347	1.425
S	0.069	0.289	0.240	.811	−0.497	0.636
C	−0.114	0.328	−0.348	.728	−0.756	0.528
B	−0.072	0.295	−0.244	.807	−0.650	0.506
ADI	−0.128	1.042	−0.123	.902	−2.170	1.915

x S	0.059	0.569	0.104	.917	−1.057	1.176
x C	0.063	0.598	0.105	.917	−1.110	1.235
x B	0.210	0.573	0.367	.714	−0.913	1.334
Others *	−1.630	0.584	−2.790	.005	−2.775	−0.485
x S	0.687	0.389	1.768	.077	−0.075	1.450
x C	−0.342	0.423	−0.810	.418	−1.171	0.486
x B *	1.079	0.393	2.748	.006	0.309	1.848
<hr/>						
Simple effect						
Baseline (ADOS)						
S	0.353	0.273	1.293	.196	−0.182	0.888
C	0.494	0.259	1.911	.056	−0.013	1.001
B *	0.536	0.234	2.289	.022	0.077	0.995
Others *	−1.091	0.411	−2.654	.008	−1.897	−0.285
S *	−0.541	0.192	−2.821	.005	−0.916	−0.165
C *	−0.791	0.228	−3.463	<.001	−1.238	−0.343
B *	0.673	0.249	2.698	.007	0.184	1.161

Note: S = social difficulty; C = communicative difficulty; B = restricted and repetitive behavior; * = significant.

Discussion

The current investigation sought to answer the highly debated question: Why is ASD less prevalent among females than among males? Addressing the question, our meta-analysis explored the variation in ASD symptom severity between females and males, in order to evaluate two different theories concerning the question of interest: The *expression hypothesis*, which postulates that the sex difference in prevalence of ASD is a result of females expressing less severe ASD traits than do males; and the *perception hypothesis*, which theorizes that the prevalence difference is a result of females' autistic expressions being perceived less as severe and better socially received, even if both sexes show similar characteristics. We developed two statistical models, one for each of the two classes of explanations, and derived competing study hypotheses for the explanations. Specifically, diagnosed females would present less severe ASD symptomatology than do their male counterparts, if the expression hypothesis is correct. By contrast, females would present more severe ASD symptomatology if the perception hypothesis is at work. To test the predictions with the highest precision, a meta-analysis was conducted on the sex disparity in symptom severity between females and males diagnosed with ASD. We also analyzed individuals' age, IQ, the diagnostic criteria of their diagnoses, and the assessment tools of their symptoms as potential moderators of the main effect of sex.

The results, which pooled effects from 103,958 historical assessment scores, indicate that, despite the publication bias found in favor of reporting males' autistic presentation relative to females', females and males with ASD show considerable similarity in ASD symptom severity. Indeed, looking closely, we discovered that females were only reported to demonstrate significantly lesser symptoms than males in terms of restricted and repetitive behavior, but not social or communicative difficulty; this effect appears to transcend IQ ranges and assessment tools used. Lastly and interestingly, our findings suggest that this pattern of sex differences may, however, be qualified by participants' age and how they were diagnosed: Males' relative restricted and repetitive behavior disappeared among adults and those who were diagnosed by ICD. For adults with ASD and those who were diagnosed by ICD, females in fact demonstrated severer overall autistic characteristics.

Expression versus Perception Hypotheses

Results showed that severity of symptoms was non-statistically distinguishable between females and males with ASD, across composite scores of ADS severity and the scores of communication and social-interaction difficulty, therefore favoring neither the expression nor the perception hypothesis. As such, we pose two competing explanations as of why these null findings may have emerged. First, both the expression hypothesis and perception hypothesis may be false; and second, they may both be true, with comparable intensity, and therefore neutralize each other in generating sex differences. Here, we lean towards the second possibility, as the expression and the perception hypothesis together form a logically comprehensive set of explanations, covering the only two directions of communication in the labeling process of ASD and leaving no room for a third option (i.e., to enable the first possibility that both hypotheses are false). Further, the two hypotheses derived were rooted in and rigorously examined by existing theories and countless empirical studies. In other words, one would need to overthrow the literature to disprove both the expression and the perception hypothesis at the same time; this work is beyond the scope of the current investigation and our imagination.

What our research implies, consequently, is that both the expression and the perception hypothesis may be true and effective to the same level, for the social and the communicative impairment of those who are younger than eighteen or were diagnosed by DSM. ASD might both be about how one demonstrates one's autistic characteristics, as well as how these characteristics are received by others. Supporting this position, our results suggest that the perception hypothesis might function as strongly as does the expression hypothesis in the domains where gender stereotyping and, thus, the perception hypothesis are more expected: social and communicative behavior. By

comparison, where gender stereotyping may not be that apparent—observable restricted and repetitive behavior—we found the expression hypothesis to outperform the perception one. Indeed, this pattern held for children younger than 18. For adults, who are likely stereotyped more than children, the data indicate that women diagnosed with ASD appear to present not only equal but actually severer ASD across symptoms than do men, bolstering the idea that the more gender stereotyping may be at work, the stronger the perception effect over the expression one.

Finally, it is intriguing to find that the relative strength of the expression and the perception hypothesis might vary between the ADOS-ADI combination and other assessment tools, as well as between the DSM and the ICD system. For the former comparison, our analysis suggests that the perception hypothesis and, thus, potential sex bias is especially strong when non-ADOS-or-ADI assessments were used. This might not be surprising and support why the two most popular assessments—ADOS and ADI—are deemed the gold standard in ASD research. However, we are surprised with the second comparison, in which DSM and ICD produced different diagnostic patterns, as they both are well constructed and widely used. Consequently, we recommend future research on how the two diagnostic systems differ in their contents as well as in the populations the systems are applied (e.g., different demographics, geographical regions etc.).

Implications

Based on our findings, we propose that the sex disparity in ASD prevalence should also be deemed a gender disparity as well. Although sex and gender often correlate with one another, the former taps into the biology that a person is born with, while the latter describes the person's social identity (for oneself) or recognition (for social others; 20). The two can consequently diverge, as in the current research where we show that the target disparity in ASD prevalence seemingly rooted in human biology may also be caused by how the disorder is socially defined and received depending on one's *social* sex—that is, gender. We did not make such a distinction between sex and gender in our analysis up front for a practical reason: It did not exist in almost all studies we found and included. The researchers either used one term (more often “sex”) without giving indication that it was not the other, applied one (in this case, gender) when it was actually the other (e.g. measured by sex), or used the two terms interchangeably. Nevertheless, as some researchers have begun advocating (3), sex and gender are two related yet separate constructs, and more attention is needed for the sex-gender analysis in ASD research in the future as we found that they may execute different effects on the diagnosis of ASD.

In addition to ASD, the models presented in the current research may apply to the analyses of other issues of demographic differences. For example, it has been documented that black Americans (cf. males) are more often incarcerated (cf. diagnosed with ASD) than are white Americans (cf. females; 21), and there is a wealth of laboratory research suggesting that explicit and implicit racism in the lab may lead to this sentencing inequality if scaled up in the real world (22). Nonetheless, it is unclear if the former really is caused by the latter, because it is difficult to study white Americans who should have been convicted but did not (cf. unsampled females), not to say they might not even exist. Given the similarity in the structure of the issues, the models used in the current research may consequently help examine whether such sentencing inequality is due to racism. Specifically, one could study whether black prisoners were given longer prison sentences than were white prisoners because of their severer transgressions (cf. symptom severity), or because of transgressions similar in nature yet received worse. If the former, the evidence would favor the expression hypothesis that black people are naturally more aggressive and thus more likely to be punished. By comparison, the latter would favor the perception hypothesis and suggest racial discrimination such that black people's behavior, although not actually different, are socially deemed more aggressive and hence more likely to be punished for.

Limitations and Future Directions

Although the current study has a number of strengths, it is not without limitations. As is true in all meta-analyses, our investigation collected no new independent sample, and can only be as good as the studies it included. As a result, it is possible that, for instance, the measures used by some of the studies we analyzed did not adequately capture the depth and range necessary to answer our question. Moreover, our models are not able to distinguish theories and mechanisms "within" each of the two classes of explanations. The work is nonetheless important and should be carried out in the future. Lastly, the chi-square distribution basing our models could be used to provide deeper insights into the nature of ASD that was beyond the scope of the current study. For instance, if females and males with ASD can be described by chi-square distributions with different degrees of freedom, as revealed by the current study's finding that supports the expression hypothesis, then one possible follow-up question may be: How many degrees of freedom exactly? The answer may provide an important clue for our understanding into the underlying dimensionality and complexity of ASD between sexes, which may in turn help us compare the many theories within the expression hypothesis.

Conclusion

Why is ASD less prevalent among females than among males? In the present paper, we first constructed two formal statistical models, one for each of both classes of exiting theories on this highly debated question; we named the two models the expression hypothesis and the perception hypothesis. Using the models, we then derived competing predictions for the expression hypothesis—females express less severe ASD traits than do males so are diagnosed less often—against the perception hypothesis—females’ autistic expressions are socially perceived as less severe so are diagnosed less often. Specifically, the models require that, for the expression hypothesis to be favored over the perception one, ASD-diagnosed males’ symptoms should be more severe than are those of diagnosed females. By contrast, diagnosed females’ symptoms should be more severe for one to favor the perception hypothesis. To test these predictions with the highest precision and the least bias, a meta-analysis was carried out using 103,958 historical assessment scores of ASD severity across ranges of age, IQ, diagnostic criteria, and assessment tools of symptoms of diagnosed individuals. The results indicate that, despite a publication bias toward males’ autistic presentation, females and males diagnosed with ASD show almost equally severe and non-conditioned ASD symptomatology. Since both our models are rooted in the literature, the evidence therefore lends strong support for both, but not neither of the expression and the perception hypothesis on the sex disparity of ASD prevalence, indicating that the diagnosis of ASD is at the same time an expression and a perception issue of the disorder.

References

1. A. P. Association, *Diagnostic and statistical manual of mental disorders (4th ed., text revision)*. (Author, Washington, DC, 2000).
2. A. P. Association, *Diagnostic and statistical manual of mental disorders (5th ed.)*. (Author, Washington, DC, 2013).
3. M.-C. Lai, M. V. Lombardo, B. Auyeung, B. Chakrabarti, S. Baron-Cohen, Sex/Gender Differences and Autism: Setting the Scene for Future Research. *Journal of the American Academy of Child & Adolescent Psychiatry* **54**, 11-24 (2015).
4. P. J. M. Van Wijngaarden-Cremers *et al.*, Gender and Age Differences in the Core Triad of Impairments in Autism Spectrum Disorders: A Systematic Review and Meta-analysis. *Journal of Autism and Developmental Disorders* **44**, 627-635 (2014).
5. E. Fombonne, Epidemiology of pervasive developmental disorders. *Pediatr Res* **65**, 591-598 (2009).
6. J. Baio *et al.*, Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill Summ* **67**, 1-23 (2018).
7. S. Begeer *et al.*, Sex Differences in the Timing of Identification Among Children and Adults with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders* **43**, 1151-1156 (2013).
8. L. Hull, W. Mandy, Protective effect or missed diagnosis? Females with autism spectrum disorder. *Future Neurology* **12**, 159-169 (2017).
9. D. H. Skuse, Imprinting, the X-Chromosome, and the Male Brain: Explaining Sex Differences in the Liability to Autism. *Pediatric Research* **47**, 9-9 (2000).
10. S. Baron-Cohen, The extreme male brain theory of autism. *Trends in Cognitive Sciences* **6**, 248-254 (2002).
11. J. N. Constantino, R. D. Todd, Autistic Traits in the General Population: A Twin Study. *Archives of General Psychiatry* **60**, 524-530 (2003).
12. A. Abu-Akel, C. Allison, S. Baron-Cohen, D. Heinke, The distribution of autistic traits across the autism spectrum: evidence for discontinuous dimensional subpopulations underlying the autism continuum. *Molecular Autism* **10**, 24 (2019).
13. K. Dworzynski, A. Ronald, P. Bolton, F. Happé, How Different Are Girls and Boys Above and Below the Diagnostic Threshold for Autism Spectrum Disorders? *Journal of the American Academy of Child & Adolescent Psychiatry* **51**, 788-797 (2012).
14. E. Giarelli *et al.*, Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and Health Journal* **3**, 107-116 (2010).
15. G. Russell, T. Ford, C. Steer, J. Golding, Identification of children with the same level of impairment as children on the autistic spectrum, and analysis of their service use. *Journal of Child Psychology and Psychiatry* **51**, 643-651 (2010).
16. L. Bartak, M. Rutter, Differences between mentally retarded and normal intelligence autistic children. *Journal of Autism & Childhood Schizophrenia* **6**, 109-120 (1976).
17. S. D. Mayes, S. L. Calhoun, Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders* **5**, 749-757 (2011).
18. T. May, K. Cornish, N. Rinehart, Does Gender Matter? A One Year Follow-up of Autistic, Attention and Anxiety Symptoms in High-Functioning Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders* **44**, 1077-1086 (2014).

19. W. Viechtbauer, Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software; Vol 1, Issue 3 (2010)*, (2010).
20. W. H. Organization. (2020, Jan.).
21. S. Bontrager, W. Bales, T. Chiricos, Race, ethnicity, threat and the labeling of convicted felons. *Criminology: An Interdisciplinary Journal* **43**, 589-622 (2005).
22. J. D. Levinson, H. Cai, D. Young, Guilty by implicit racial bias: The guilty/not guilty Implicit Association Test. *Ohio State Journal of Criminal Law* **8**, 187 (2010).