HHS Public Access

Author manuscript

Psychol Bull. Author manuscript; available in PMC 2022 March 29.

Published in final edited form as:

Psychol Bull. 2021 March ; 147(3): 268–292. doi:10.1037/bul0000310.

A Quantitative Meta-Analysis of Face Recognition Deficits in Autism: 40 Years of Research

Jason W. Griffin, M.A.,

Department of Psychology, Pennsylvania State University, 423 Moore Building, University Park, PA, 16802

Russell Bauer, B.A.,

Department of Psychology, Pennsylvania State University, 423 Moore Building, University Park, PA, 16802,

K. Suzanne Scherf, Ph.D. [Assoc. Professor Psychology]

Pennsylvania State University, 113 Moore Building, University Park, PA, 16802

Abstract

The ability to recognize an individual face is essential to human social interaction. Even subtle errors in this process can have huge implications for the way we relate to social partners. Since autism spectrum disorder (ASD) is characterized by deficits in social interaction, researchers have theorized about the potential role of atypical face identity processing to the symptom profile of ASD for more than 40 years. We conducted an empirical meta-analysis of this large literature to determine whether and to what extent face identity processing is atypical in ASD compared to typically developing (TD) individuals. We also tested the hypotheses that the deficit is selective to face identity recognition, not perception, and that methodological variation across studies moderates the magnitude of the estimated deficit. We identified 112 studies (5,390 participants) that generated 172 effect sizes from both recognition ($k = 119$) and discrimination ($k = 53$) paradigms. We employed state-of-the-art approaches for assessing the validity and robustness of the analyses. We found comparable and large deficits in ASD for both face identity recognition (Hedge's $g = -0.86$) and discrimination (Hedge's $g = -0.82$). This means that the score of an average ASD individual is nearly 1 SD below the average TD individual on tasks assessing both aspects of face identity processing. These deficits generalize across age groups, sex, IQ scores, and task paradigms. These findings suggest that deficits in face identity processing may represent a core deficit in ASD.

Keywords

systematic review; ASD; face identity; face memory; face perception

Corresponding author Correspondence concerning this article should be addressed to K. Suzanne Scherf, Ph.D., Assoc. Professor of Psychology, Pennsylvania State University, 113 Moore Building, University Park, PA, 16802, (814)-867-2921, suzyscherf@psu.edu.

Data and Code Availability

Our data and code are available through an Open Science Framework repository (available at https://osf.io/tcpmg/.)

Faces are a primary source of information for identifying people as unique individuals, particularly from a distance. The need to recognize individuals is critical for social interaction. The ability to recognize people enables us to determine whether they are personally familiar to us (e.g., this is my boss), predict their behavior (e.g., she will ask me not to be late), and shape our own behavioral responses to them (e.g., I will listen respectfully). This recognition process is complex and requires an integrated network of brain regions that extend into all four lobes of the brain (Elbich & Scherf, 2017). Although there are vast individual differences in face recognition ability (e.g., Elbich & Scherf, 2017), typically developing (TD) adults can recognize upwards of 5,000 individual faces (Jenkins, Dowsett, & Burton, 2018), across impressive variations in visual information (e.g., lighting, viewpoint) and context (e.g., emotional expressions, age, occlusion by paraphernalia). Although there are multiple important cues that facilitate person identification (e.g., voice, body, contextual information); under many conditions, information from the face is disproportionately weighted in this process (see Burton, Wilson, Cowan, & Bruce, 1999; O'Toole et al., 2011; Rice, Phillips, Natu, An, & O'Toole, 2013).

Errors in face recognition, even when subtle, can have huge implications for social interaction. A person who fails to recognize their boss in the above example may lose out on the opportunity to learn from the situation (e.g., that their work behavior needs to change), which is likely to have important consequences for the relationship. Similarly, a person who mistakes the identity of their boss for that of their friend may chose a very different behavioral response that is only appropriate for a social interaction with the friend (e.g., "I was only 5 minutes late!") and that could damage the relationship. Appropriate and adaptive social interaction is contingent upon the ability to identify our social partners as unique individuals. Face recognition provides one of the fastest ways of doing so because it relies on vision, which is a distant sense. When this process is erroneous or inconsistent, there are consequences for social interactions. Therefore, studying potential disruptions in face recognition abilities has important implications for understanding difficulties in many aspects of human social interaction.

Autism spectrum disorder (ASD) is a lifelong neurodevelopmental disorder that is characterized by difficulties in human social interaction. It is diagnosed on the basis of behavioral symptoms in two domains: social communication and restricted, repetitive patterns of behaviors, interests, or activities (American Psychiatric Association, 2013). Importantly, all of the impaired behaviors identified in the social communication domain are directly or indirectly reliant on face processing skills. For example, individuals with ASD often exhibit abnormal social approach and sometimes a complete inability to initiate social interactions, both of which likely contribute to difficulties developing and maintaining friendships and adjusting behavior to social contexts. These behaviors could stem, in part, from an inability to recognize people, particularly by their face. Although face identity recognition has never been a criterion used to diagnose ASD, researchers have theorized about the potential importance of atypical face recognition abilities to the behavioral symptom profile of ASD for more than 40 years. This has generated hundreds of empirical articles investigating face recognition abilities in ASD. However, this literature is rife with inconsistent findings, which was noted in two prior summary reviews (Tang et al., 2015; Weigelt et al., 2012). To date, there are no *quantitative* meta-analyses of this research to

determine whether it indicates deficits in face recognition in ASD and if so, how large is the effect. Here, we present the results from a quantitative meta-analysis of the face identity processing literature comparing ASD and TD individuals.

The paper is organized as follows. First, we provide a general overview of the literature investigating face identity processing deficits in ASD and describe some of the issues that have likely contributed to inconsistent findings. Then we describe the conclusions from the existing summary reviews of this literature. Finally, we present the methodology and findings of our quantitative meta-analysis of this literature.

Reviewing the existing literature

Although deficits in face identity processing are not a diagnostic feature of ASD, atypical face identity processing is one of the most researched behaviors in ASD. The first study of face identity processing in ASD empirically evaluated the face recognition abilities of children with ASD in 1978 (Langdell). In this study, participants viewed photographs of peer faces and face identity recognition was tested on the basis of isolated face features (i.e., nose, eyes) in lower and upper face halves. Langdell reported that children with ASD were generally worse at recognizing faces compared to age-matched TD children except when using the lower half, compared to the upper half, of the face (1978). Since this early study, hundreds of studies have been reported, making face identity processing one of the most studied topics in ASD. As we reviewed this literature, it became clear that there were several methodological factors that varied across studies, which may have contributed to heterogeneity in findings. These include the sample size, empirical strategies for verifying an ASD diagnosis, age range of participants, distribution of participant sex, and the range of paradigms used to measure face identity processing. Next, we provide a brief overview of these different methodological factors.

Inconsistent findings about face identity processing deficits in ASD may be related to variability in sample size. Before there was much public awareness about ASD and it became a priority for federal funding in the United States, there were limited resources for recruiting large samples of participants with ASD into research studies. For example, some early studies only included seven participants with ASD (see Gepner, de Gelder, & de Schonen, 1996), whereas more recent studies have as many as 140 participants with ASD (see Oerlemans et al., 2013). This variability in sample size is important because small sample sizes are less representative of the population, generate more biased effect sizes, and are more prone to making inferences in the wrong direction (Gelman & Carlin, 2014).

There is also considerable variability in how studies have empirically confirmed ASD diagnoses. For example, clinicians typically diagnose ASD using the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013) or *International and Statistical Classification of Diseases and Related Health Problems* (ICD-10; World Health Organization, 2004). In contrast, the gold standard for assessing and confirming an ASD diagnosis in research studies is through the use of empirically validated diagnostic interviews, such as the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) and/or the Autism Diagnostic Interview (ADI-R; Lord et al., 1994). The training

to conduct these interviews is expensive and not offered very frequently. As a result, many of the existing studies rely on community diagnoses and/or employ surveys of autism behaviors such as the Autism-Spectrum Quotient (AQ; Baron-Cohen et al., 2001) or Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003). These measures have empirically validated cut-off scores that are associated with an ASD diagnosis but are not diagnostic tools in and of themselves. As a result, variability in the assessment of ASD diagnoses across studies could contribute to heterogeneity in findings if (1) the assessments are differentially sensitive to symptom severity and (2) symptom severity and face identity processing are related. Indeed, there is evidence that both adolescents with ASD (Scherf, Elbich, Minshew, & Behrmann, 2015) and adults with more autism-like personality traits (Halliday, MacDonald, Scherf, & Tanaka, 2014) exhibit worse face identity performance as a function of the magnitude of their symptom severity or traits.

Studies also vary considerably in the age range of the participants. For example, some studies include individuals across a wide age range like 16–53 years (Williams, Goldstein, & Minshew, 2005) or 15–42 years (McPartland, Dawson, Webb, Panagiotides, & Carver, 2004), while other studies include individuals from a narrower age range like 6–10 years (Robel et al., 2004) or 13–17 years (O'Hearn et al., 2014). This is important to consider because of age-related changes in face identity processing among TD individuals (e.g., Germine, Duchaine, & Nakayama, 2011) and evidence that there may be a plateau in the development of face processing skills in adolescents with ASD (see Picci & Scherf, 2014). Therefore, variability in participant age may also contribute to the inconsistency in findings across studies because differences between ASD and TD groups may be influenced by the age of the sample. For example, studies with younger samples might find small or no group differences in face identity processing abilities, whereas those with adolescent and/or adult samples might find the largest group differences.

The make-up of participant groups varies by study as well. There is a strong male bias in the current status of ASD diagnoses (4:1 male to female). As a consequence, most studies include samples that reflect this asymmetrical sex distribution. However, there is a growing awareness that the ASD phenotype may manifest differently in males and females (e.g., Lai et al., 2017b), particularly in terms of face processing behavior (see Whyte & Scherf, 2017). As a result, studies have begun recruiting a more equal number of male and female participants with ASD. There are several studies in the face identity processing literature that reflect this approach (e.g., Arkush, Smith-Collins, Fiorentini, & Skuse, 2013; Ewing et al., 2018; Philip et al., 2010; Reinvall, Voutilaninen, Kujala, & Korkman, 2013; White, Hill, Winston, & Frith, 2006). If face identity processing is differentially impaired in males and females with ASD, then variations in the distribution of participant sex could explain inconsistencies in findings of group differences across studies as well.

Finally, researchers have used many paradigms to investigate face identity processing. Paradigms generally measure either *face identity recognition* or *face identity discrimination*. Recognition paradigms require participants to invoke a mental representation of face identity in the absence of the percept (i.e., as in over a delay period). They include "Old/New" recognition, "n-back", and standardized neuropsychological tests such as the Cambridge Face Memory Task (CFMT; Duchaine & Nakayama, 2006), Wechsler Memory

Scale (WMS; Weschsler, 1997), and the Developmental NEuroPSYchological Assessment (NEPSY; Korkman, Kirk, & Kemp, 2007). Discrimination paradigms require participants to perceptually discriminate between multiple faces on the basis of identity using available percepts (i.e., simultaneous presentation). These paradigms include Matching-to-sample, "Same/different" judgments, sorting based on identity, and the Benton Face Recognition Test (BFRT; Benton, Sivan, Hamsher, Varney, & Spreen, 1983). There are some reports indicating that these two face identity behaviors are correlated (e.g., Bruce et al., 2018; Fysh, 2018; Verhallen et al., 2017). However, variability in findings across studies could reflect the degree to which any particular paradigm primarily tests face identity recognition or discrimination abilities.

Is there consensus about whether there are there face identity processing deficits in ASD?

To date, there are two summary review articles of this large literature (Tang et al., 2015; Weigelt et al., 2012). Both reviews computed a frequency score of the number of studies with significant findings in which individuals with ASD performed worse than TD individuals on measures of face identity processing. Interestingly, perhaps because of the size, breadth, and extent of variation in findings/paradigms across studies, the two reviews provided different conclusions about face identity processing in ASD.

The initial review by Weigelt et al. (2012) summarized a total of 90 experiments across 33 years of research. This included 28 studies investigating markers of face expertise (e.g., face inversion effect), 14 studies of face identity recognition, and 24 studies of face identity discrimination among individuals with ASD. The researchers concluded that individuals with ASD exhibit all the same behavioral markers of expert visuoperceptual processing of faces (i.e., face inversion effect) as do TD individuals. They interpreted this finding to mean that there are no *qualitative* differences in face identity processing in ASD. In contrast, they reported that studies of face identity recognition, as defined by those requiring participants to maintain a mental representation of the face for more than 30 seconds, largely reported a deficit in ASD. They reported that studies using face identity discrimination tasks do not consistently report ASD deficits. As a result, Weigelt et al. (2012) proposed that as memory demands increase, face recognition behavior is increasingly impaired, but that face identity perception is intact, in ASD.

Subsequently, Tang et al. (2015) conducted a summary review of 25 studies that were either published following the Weigelt et al. (2012) review or that were not included in this earlier review. These studies included both visuoperceptual expertise and face identity recognition studies. In contrast to Weigelt et al. (2012), Tang et al. (2015) reported consistent deficits in the qualitative components of face identity processing in ASD (e.g., face inversion effect). They also found that the majority of studies (71%) they reviewed reported ASD deficits in face identity recognition behavior. However, they also noted that nearly 1/3 of the studies failed to support this finding and suggested that methodological heterogeneity likely contributed to inconsistent findings.

Our goal was to build upon these two prior reviews of face identity processing in ASD by including all relevant effect sizes reported that meet our inclusion criteria (see Method). This allowed us to include both sets of data reviewed separately in previous studies together with the newest research. We used a quantitative strategy to assess the presence, magnitude, and external validity of face identity processing deficits in ASD. We also tested both the predictions that were identified in these earlier studies. First, we evaluated the potential contribution of each of the methodological factors we discussed above to determine whether they influence the pattern of group differences in face identity recognition and discrimination behavior, which is a test of the hypothesis proposed in the Tang et al. (2015) review. Second, we evaluated whether there is a selective or disproportionate deficit in ASD for face identity recognition compared to face identity discrimination. This is a direct test of the hypothesis proposed in the Weigelt et al. (2012) review.

The Present Study

Although there have been quantitative meta-analyses investigating the effect size of multiple deficit behaviors in ASD, including social visual attention (Chita-Tegmark, 2016; Frazier et al., 2017), social cognition (Velikonja, Fett, & Velhorst, 2019), facial emotion recognition (Uljarevic & Hamilton, 2013), working memory (Wang et al., 2017), and executive functioning (Lai et al., 2017a), there is no such meta-analysis of face identity processing deficits in ASD. Therefore, the existing literature requires a comprehensive quantitative review of the findings from the last four decades of empirical research on face identity recognition and discrimination behavior in ASD.

Empirical meta-analyses pool findings across and within multiple studies using a principled approach for estimating a summary effect size. Meta-analyses also provide the opportunity to systematically investigate the extent to which such an effect is externally valid as well as the impact of potential moderating variables on this effect. Here, in this meta-analysis of the ASD face identity processing literature, we had five primary goals: (1) determine if there are reliable deficits in face identity recognition behavior in ASD compared to TD individuals; (2) if so, precisely estimate the magnitude of these deficits in a summary effect size across studies; (3) test the hypothesis that potential moderating variables such as participant age, sex, and IQ, as well as task paradigm, and study design and reporting quality influence the estimated magnitude of these deficits; (4) test the hypothesis that face identity processing deficits in ASD are specific to face identity recognition not face identity discrimination; and (5) if deficits exist in identity discrimination as well, evaluate the magnitude of the deficit and whether methodological factors influence it.

Method

We followed all PRISMA guidelines (see Supplemental Materials) and include a detailed flow diagram (Figure 1) of our study screening and identification process (Moher, Liberati, Tetzlaff, & Altman, 2009).

Study Search and Identification

The diagnostic criteria, and thus nomenclature, of autism changed considerably over the last 40 years when this literature was published. Therefore, we included study search terms that would capture the full range of terminology used to classify individuals on the autism spectrum (particularly by the multiple versions of the DSM) across this time period time.

We used three sources for initially identifying relevant articles. First, on February 11, 2019, we conducted a literature search in the PubMed database with the following search parameters: (autis* OR asperger) AND (face OR facial) AND (recall OR memory OR recognition OR identity). Second, we added the references from Weigelt et al. (2012) and Tang et al. (2015), the reviews of face identity processing in ASD, to our initial search results. We limited the search to experimental studies that employed tasks of behavioral face identity processing, including both recognition and discrimination tasks comparing ASD and TD participants. Neuroimaging, eye tracking, and electroencephalographic studies were also reviewed in case they also collected behavioral outcome measures using face identity recognition and/or discrimination tasks in the context of the neuroimaging protocol or in addition to the protocol. We searched for studies that reported a measure of behavioral accuracy, which includes either number correct, percentage correct, number of errors, or percentage of errors. Third, we conducted a "manual reference search" by screening the reference lists of each study that met full inclusion criteria from the original search results (i.e., PubMed search).

Inclusion and Exclusion Criteria

We employed the following inclusion criteria: articles must (1) report using a task of unfamiliar face identity processing in either a recognition or discrimination paradigm; (2) include an ASD group; (3) include a TD control group; (4) include behavioral accuracy scores for both ASD and TD groups; and (5) report a between groups effect size (e.g., Cohen's *d)* or information to compute one (e.g., t, M, SD, N, se). Studies were included as long as ASD groups were characterized as having a diagnosis of autism, ASD, Asperger syndrome, autistic disorder, or pervasive developmental disorder – not otherwise specified (PDD-NOS).

Studies were excluded if they were not (1) published in English or (2) peer-reviewed. If articles met inclusion criteria, but did not report enough information in the published manuscript to compute an effect size, we emailed corresponding authors to request the relevant data.

Power Analysis

Statistical power is rarely considered when planning a meta-analysis, likely because software to compute statistical power for meta-analyses is not readily available (but see Griffin, 2020). However, Valentine, Pigott, and Rothstein (2009) provide the set of formulas for computing statistical power for meta-analyses, which only require four anticipated input parameters: 1) total number of effect sizes; 2) average sample size per group; 3) between-study heterogeneity (variance), and 4) summary effect size. Power is computed with the following formula:

$$p = 1 - \Phi(c_\alpha - \lambda^*) \quad (1)$$

where c_α represents the critical value (e.g., $\alpha = .05$) and λ^* represents the mean of a normal Z distribution when the overall effect is statistically different from 0. To obtain λ^*, one needs to compute the anticipated sampling variance of a typical study (see Equation 4) and sampling variance of the random effects. The random effects variance is computed by adding the typical sampling variance and the heterogeneity estimate (τ^2). Since meta-analysis is a weighted technique, the random effects variance is then weighted by number of studies. With the weighted variance computed, we can compute λ^* with the following:

$$\lambda^* = \frac{\overline{d} - 0}{\sqrt{v}} \quad (2)$$

Based on the two previous review papers, we anticipated that 50 studies would meet inclusion criteria with an average participant group size of 20 and large heterogeneity between studies (.33 = small, 1 = moderate, 3 = large). Overall, we anticipated either a small ($d = .25$) or medium effect size ($d = 0.5$). The calculated power estimate was 79.5% to detect a small effect size and 99.8% to detect a medium effect size. Regardless, we planned to incorporate as many studies as possible to comprehensively represent the current state of the literature.

Study Selection

The full study search, identification, and selection process is shown in Figure 1. We applied our inclusion criteria in two steps. First, we screened the titles and abstracts of all articles returned from the initial PubMed search. Since this was the first level of screening, we emphasized overinclusion to maximize yield. For example, abstracts were only rejected on the basis of exclusion criteria (i.e., not published in English, not an empirical paper, no ASD vs TD group comparison). After this first set of articles was pared down, we assessed the remaining articles for inclusion by evaluating each full text article for group characteristics (i.e., ASD, TD), task paradigm (e.g., CFMT), and effect sizes (*M*, *SD*, *N*, *t*, *d*).

Researchers frequently conduct a “manual reference search” by screening the reference lists of all included studies from the initial search. However, the written description of this approach is minimal at best. We find this norm in meta-analytic reviews to be incredibly difficult to replicate as there is no systematic procedure for how articles are “manually screened”. This is not a trivial task. For example, our database search yielded 105 articles that contained 5,143 references to be “manually screened.” This manual reference approach typically generates numerous duplicates and irrelevant studies to the meta-analysis.

To address these issues, we developed a novel and systematic procedure for extracting reference list titles that is transparent and reproducible. First, we copied the reference lists from each article into a text editor. Next, we used *RegEx*, a pattern recognition language that detects patterns in strings of text, that is available in all text editing software to extract the article titles (see Supplemental Figure 1). Using this approach, we generated a database of reference titles, removed duplicates, and proceeded with the general strategy for screening

articles (see above). We extracted another seven studies from this approach to include in the analysis.

JG and RB independently reviewed all studies for inclusion. After title and abstract screening, all the studies identified from this initial review from both authors were evaluated for inclusion and exclusion criteria in a full text review. Any criterion discrepancies were discussed by JG and RB until consensus was reached. If there was no consensus, KSS helped resolve the disagreement about study inclusion. The process of selection is illustrated in Figure 1 and the full set of studies included in the analysis is reported in Table 1.

Data Extraction

Standardizing effect sizes.—We extracted information to compute the standardized mean difference, Cohen's *d*, including means, standard deviations, standard errors, sample sizes, and test statistics from each study. When raw data were provided, we calculated the means, standard deviations, and sample size manually. Many studies reported multiple effect sizes (e.g., from multiple face identity tasks or condition comparisons). As a result, the data were inherently hierarchical, such that accuracy scores were nested within study. Each study and unique effect size were coded to account for this multilevel structure. Cohen's *d* is known to overestimate the population effect size, especially when degrees of freedom are low ($df < 50$), which is characteristic of many of the studies in this sample. Therefore, we converted Cohen's *d* scores to the bias-corrected Hedge's *g* for meta-analytic synthesis (Cumming, 2012; Hedges, 1981; Viechtbauer, 2010).

For means, standard deviations, and sample sizes, we used the 'metafor' package in R (Viechtbauer, 2010) to compute Cohen's *d*. For *t* statistics, we used the following set of formulas for conversion to Cohen's *d* (Viechtbauer, 2010; Cumming, 2012):

$$d = t \times \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \tag{3}$$

where N_1 and N_2 are the sample sizes of ASD and TD groups. Sampling variance was defined and computed with:

$$V_i = \frac{N_{1i} + N_{2i}}{N_{1i} N_{2i}} + \frac{d_i^2}{2(N_{1i} + N_{2i})} \tag{4}$$

Where N_{1i}, N_{2i}, and d_i are the study specific sample sizes and effect sizes. Finally, we converted Cohen's *d* scores to Hedge's *g* with the following:

$$\text{Hedge}'s\ g = \left(1 - \frac{3}{4\,df - 1}\right) \times d \tag{5}$$

Moderator Variables.—To investigate the potential effects of methodological variation on the summary effect sizes, we extracted data on participant age in years (*M*, *SD*), sex (% male), and Full Scale IQ (*M*, *SD*) for both ASD and TD groups when available. We also obtained information regarding the strategy or diagnostic system researchers used to

confirm the diagnosis of autism (e.g., ADOS, ICD-10, DSM-IV), and measures of symptom severity in the ASD group. Finally, we extracted information about the specific type of face processing task (e.g., Old-New Recognition task, CFMT, Match-to-Sample) that was used to test either face identity recognition or discrimination abilities.

Design and Reporting Quality Index (DRQI).—Consistent with previous meta-analytic approaches (e.g., Frazier et al., 2017; Tang et al., 2015), we developed a metric to quantify the quality of the study design and methodological reporting of each study. To do so, we created a Data Reporting Quality Index (DRQI) that we adapted from the literature (see Frazier et al., 2017; Kmet, et al., 2004) and used to score each study. The index ranged from 0–15 points with higher scores indicating better design and reporting quality. The DRQI includes components that evaluate the quality of reporting (e.g., participant demographics) and of study design (e.g., group matching features). For example, studies that matched participant groups (ASD and TD) on demographic variables such as age, IQ, and sex, were coded with a 1 if the group means were matched and a 2 if participants were matched at the individual level. The studies were coded with a 0 if there was no matching on these variables. Regarding the use of a strategy to confirm the autism diagnosis, the scale ranged from 0–3. Studies that did not confirm the diagnosis were scored with a 0; studies that employed a measure of autism-like symptomology (e.g., SCQ) were scored a 1; studies using the DSM or ICD-10 to evaluate the diagnosis were scored a 2; and, studies using the ADOS or ADI were scored the maximum value of 3. Supplementary Table 3 illustrates the DRQI component and total scores for each study.

Statistical Analysis

We used the ‘metafor’ package for the statistical software program R (R Core Team, 2018; Viechtbauer, 2010) to analyze the data. To address our first and second goals, we estimated a summary effect size for deficits in face identity recognition behavior in ASD using multilevel random-effects modeling. This approach accounts for hierarchical structure and dependencies in observations (i.e., multiple effect sizes from one study) and in so doing, achieves higher statistical power by maximizing the total number of effect sizes (Assink & Wibbelink, 2016; Cheung, 2014; 2019). To do so, we fit a three-level meta-analytic model to partition the following sources of variance: sampling variance of the observed effect sizes (Level 1), between study variability (Level 2), and between study variation within effect sizes (Level 3; Assink & Wibbelink, 2016). This approach also allowed us to evaluate whether variation in the summary effect size could be accounted for by methodological variations across studies (i.e., design or sample characteristics).

Finally, to test the hypothesis offered by Weigelt et al. (2012), that face identity recognition, not face identity discrimination, predicts the magnitude of face identity processing deficits in ASD, we compared the summary effect sizes for face identity recognition ($k = 119$) and face identity discrimination ($k = 53$) behaviors. To compare these sets of studies using meta-analysis, we incorporated both face identity recognition and face identity discrimination effect sizes into a multivariate multilevel model to account for the potential correlation between these two tasks. Since this model evaluates two potentially distinct, but related, outcomes (recognition and discrimination on the basis of identity), one must

specify a correlation between these outcomes in the model (Viechtbauer, 2010). Among TD individuals, the correlation between these kinds of face identity processing tasks is approximately 0.3 – 0.5 (Bruce et al., 2018; Fysh, 2018; Verhallen et al., 2017). Among individuals with ASD, this correlation is rarely reported. Therefore, we modeled this correlation with multiple imputations ($r = 0.3$. $r = 0.6$, $r = 0.9$) to determine whether it impacted the findings.

Sensitivity to small-study effects.—The validity of meta-analysis estimates can be compromised by small-study effects, which is the phenomenon that smaller studies can show exaggerated estimates compared to large studies. One of the most well-known causes of small-study effects is publication bias, which reflects the fact that studies are generally published based on the prominent factor of statistical significance instead of multiple factors like statistical power, magnitude of the effect, and overall quality of the study. Funnel plots, which quantify symmetry, are a standard method for identifying the presence and impact of small-study effects on traditional, univariate meta-analyses. They display the effect size estimates from individual studies as a function of precision, which is defined as either the study-specific sample size, variance, standard error, inverse variance, or inverse standard error. In so doing, small-study effects can be identified as those that have magnified effect sizes that are relatively distant in an asymmetrical way from the summary effect. Quantitative techniques such as Eggers' Regression test have been developed to test for funnel plot asymmetry; however, these methods are not appropriate for hierarchical data (Egger, Smith, Schneider, & Minder, 1997).

To evaluate the robustness of our multilevel, meta-analyses to small-study effects, we assessed funnel plot asymmetry based on the recommendations of Nakagawa and Santos (2012). Specifically, we included a measure of precision (i.e., standard error) as a predictor in our multilevel model, which manages statistical dependency of multiple within-study effect sizes and is conceptually identical to the Egger's regression test (i.e., a standard method to quantify the funnel plot asymmetry). After determining the small-study influential effect sizes, we computed the summary effect size without those influential cases. Separately, we evaluated the sensitivity of the summary effect sizes from recognition and discrimination studies to small study effects.

Evaluation of methodological factors.—After estimating the summary effect size without the influence of small-study effects, we extended the three-level meta-analytic model to include the continuous moderator variables of age, sex (percentage male), full scale IQ, and DRQI scores and the categorical moderator of task paradigm. As recommended, we started by evaluating each moderator as a predictor of the effect size one at a time to evaluate the unique influence on the summary effect independent of the other variables (Assink & Wibbelink, 2016; Cheung, 2014; Viechtbauer, 2010). Additionally, since not all studies reported information on all moderator variables, this strategy achieves maximum statistical power for each variable. To investigate the effect of task paradigm, we only included paradigms for which there was a substantive cluster ($k \geq 5$). We evaluated the potential influence of these moderating variables on the summary effect sizes from the recognition and discrimination studies separately.

Results

Study Selection

In total, we included 172 effect sizes from 112 unique empirical articles in the analysis (indicated by an asterisk in References). The final dataset included 5,390 individual participants (ASD = 2,612, TD = 2,778) whose average age ranged from 4- to 44-years-old. Specifically, we included 119 face identity recognition and 53 face identity discrimination effect sizes in the analyses. The full process of study identification, screening, and selection is documented in Figure 1. The sample sizes, study characteristics, participant demographics, and face recognition/discrimination paradigm and the contributing effect size for each study are reported in Tables 1–2.

Is There a Deficit in Face Identity Recognition Behavior in ASD?

The first analysis evaluated whether there is a deficit in face identity recognition behavior in ASD, and if so, provided an estimate of the magnitude of the deficit. To do this analysis, we submitted the 119 effect sizes from the recognition tests to the multilevel model with group (ASD, TD) as a fixed factor. This analysis revealed a large summary effect size (Hedge's g = −0.86, 95% CI [−1.00, −0.72], $p < .0001$), indicating prominent face identity recognition deficits in ASD compared to TD individuals (see Figure 2).

Sensitivity to small-study effects.—We evaluated the influence of small-study effects on the overall summary effect size by using a modified Egger's Regression test to quantify funnel plot asymmetry (Nakagawa & Santos, 2012). There was significant funnel plot asymmetry (b = −4.50, se = 0.53, 95%CI [−5.53, −3.47], $p < .0001$; see Figure 4a), indicating the potential presence of small-study effects. After the removal of extreme effect sizes, we no longer observed funnel plot asymmetry (b = −1.24, se = 0.64, 95%CI [-2.49, 0.00], $p > .05$; see Figure 4b). After removal of these extreme effect sizes (n = 7), there was still a large summary effect size estimate (Hedge's g = −0.76, se = 0.05, 95% CI [−0.86, −0.67], $p < .0001$) indicating deficits in face recognition in ASD. This finding suggests that our summary effect size was robust to different sources of bias related to funnel plot asymmetry.

Do Methodological Factors Influence the Magnitude of this Face Recognition Deficit?

Consistent with our choice in using a random-effects model, we observed considerable heterogeneity among the included effect sizes (Q = 473.04, $p < .0001$, I^2 = 77.8%). Specifically, the model revealed that 22.2% of the total variance was attributed to sampling variability (Level 1), 70.3% of the total variance was attributed to between-study variability (Level 2), and 7.5% of the total variance was attributed to between study variation within effect sizes (Level 3). This observed degree of between-study heterogeneity exceeds recommendations for proceeding with subsequent moderator analyses to evaluate the potential sources of this heterogeneity (i.e., $I^2 > 75\%$; Hunter & Schmidt, 1990).

In contrast to predictions from Tang et al. (2015), the summary effect size for face identity recognition was not significantly moderated by differences in participant age, $F(1, 113)$ = 1.65, b = −0.01, se = 0.01, 95% CI[−0.03, 0.01], p = .20, full scale IQ, $F(1, 70)$ = 1.17,

b = −0.01, se = 0.01, 95% CI[−0.03, 0.01], p = .28, sex, F(1, 102) = 0.65, b = 0.50, se = 0.62, 95% CI[−0.73, 1.73], p = .42, DRQI score, F(1, 117) = 0.72, b = 0.03, se = 0.04, 95% CI[−0.04, 0.10], p = .40, or recognition paradigm, F(4, 104) = 1.77, p = .78. These findings indicate that variation in these methodological factors did not consistently influence the magnitude of face identity recognition deficits in ASD in the studies reviewed here.

Given the limitations of behavioral paradigms evaluating face identity processing used as participants undergo neuroimaging or eye-tracking, we re-evaluated the summary effect size characterizing face identity recognition deficits in ASD after excluding these effect sizes (k = 31; see Table 1). Without these effect sizes, the summary effect size remained large and statistically significant (Hedge's g = −0.74, se = 0.05, 95% CI [−0.84, −0.65], p < .0001).

Is the Deficit Specific to Face Identity Recognition Behavior?

Finally, we tested the hypothesis that face identity processing deficits are fundamentally related to memory, and not perceptual, processes (see Weigelt et al., 2012). To do this analysis, we submitted all 172 effect sizes to a multivariate multilevel model with group (ASD, TD) as a fixed factor and process (recognition, discrimination) as a moderating factor. We found no moderating effect of the face identity process (b = −0.13, se = 0.11, 95% CI [−0.35, 0.10], p = .27). Specifically, there were comparable deficits in performance on face identity recognition and face identity discrimination tasks among the ASD participants. Sensitivity analyses indicated that this was true for each correlation imputation (r = 0.6: b = −0.13, se = 0.11, 95% CI [−0.35, 0.09], p = .26; r = 0.9, b = −0.13, se = 0.11, 95% CI [−0.35, 0.08], p = .23). To follow up, we estimated the unique summary effect size indicating deficits in face identity discrimination behavior (k = 53), which also revealed a large summary effect size, (Hedge's g = −0.82, se = 0.12, 95% CI [−1.04, −0.59], p < .0001; see Figure 3).

We also evaluated bias from potential small-study effects on the magnitude of this summary effect. There was significant funnel plot asymmetry for face identity discrimination effect sizes (b = −5.45, se = 0.68, 95%CI [−6.79, −4.10], p < .0001; see Figure 4c), indicating the potential presence of small-study effects. After the removal of extreme effect sizes (n = 6), we no longer observed funnel plot asymmetry (b = −0.95, se = 0.92, 95%CI [−2.75, 0.85], p = .30) and we still observed a large summary effect size estimate (Hedge's g = −0.65, se = 0.07, 95% CI [−0.78, −0.52], p < .0001; see Figure 4d). These findings indicate that the summary effect size describing deficits in face identity discrimination behavior was also robust to different sources of bias related to funnel plot asymmetry.

Do Methodological Factors Influence the Magnitude of the Face Identity Discrimination Deficit?

Given the finding that there are consistent deficits in face identity discrimination in ASD across studies, we also investigated the influence of methodological variation on the summary effect size characterizing the magnitude of this deficit. The summary effect size was not significantly moderated by differences in participant age, F(1, 50) = 1.54, b = −0.02, se = 0.01, 95% CI[−0.05, 0.01], p = .22, full scale IQ, F(1, 19) = 0.92, b = −0.02, se = 0.02, 95% CI[−0.07, 0.03], p = .35, sex, F(1, 44) = 1.38, b = 1.39, se = 1.19, 95% CI[−1.00,

3.80], $p = .25$, DRQI score, $F(1, 51) = 0.04$, $b = 0.01$, $se = 0.04$, 95% CI[−0.08, 0.09], $p = .85$, or discrimination task, $F(3, 49) = 0.31$, $p = .95$. Finally, we also excluded effect sizes obtained while participants were undergoing neuroimaging or eye-tracking ($k = 3$; see Table 2) and the summary effect size was unaffected (Hedge's $g = −0.82$, $se = 0.13$, 95% CI [−1.07, −0.57], $p < .0001$). These findings indicate that in spite of the methodological variation across studies, the magnitude of the deficit in face identity discrimination behavior is not impacted.

Discussion

ASD is characterized by deficits in human social interaction and researchers are trying to understand the underlying mechanisms contributing to these deficits. Researchers have theorized about the contribution of atypical face identity processing to the behavioral symptom profile of ASD for more than 40 years. This is because so many of the social communicative difficulties in ASD are directly or indirectly reliant on face processing skills, like face recognition. For example, face recognition is essential for determining whether a social partner is personally familiar to us, how to predict their behavior, and how to select our own behavioral responses to them. It is possible that even subtle deficits in this fundamental social cognitive skill will have downstream consequences for multiple aspects of social communication.

Here, we present a quantitative meta-analysis of this body of work to evaluate the consistency and magnitude of differences in face identity processing between ASD and TD individuals. We used rigorous approaches for identifying effect sizes to be included in the meta-analysis and for performing the analyses. First, we established clear definitions of face identity processes. We operationalized face identity recognition as the ability to recognize a face by invoking a mental representation of face identity in the absence of the percept (i.e., as in over a delay period). We operationalized face identity discrimination as the ability to perceptually discriminate between multiple faces on the basis of identity using available percepts (i.e., simultaneous presentation). Operationalizing these two kinds of face identity processing tasks allowed us to test a hypothesis from an existing review article (Weigelt et al., 2012), that proposed the face identity processing deficits in ASD are selective for recognition, and do not extend to discrimination. Our approach identified 172 effect sizes (119 from recognition and 53 discrimination paradigms) that represented 5,390 unique participants. In our analysis, the studies were weighted as a function of sample size and we used multilevel modeling to estimate the summary effect size, which accounts for hierarchical structure and dependencies in effect sizes (i.e., multiple effect sizes from one study). This approach maximized both the statistical power and robustness of our results. Finally, we evaluated how vulnerable the results were to small-study effects arising from publication bias, poor study quality, or chance.

Consistency in Face Recognition Deficits in ASD Across Studies

This meta-analysis revealed large face identity recognition deficits in ASD compared to TD groups (Hedge's $g = −0.86$). Essentially, this means that the score of an average ASD individual is almost 1 SD below the average TD individual on face identity recognition

tasks. Put another way, an average ASD person will score below ~81% of TD individuals on face recognition tasks. In addition, we found that this effect persists across a wide array of task paradigms, age groups, IQ scores, sex distributions, and design and reporting quality of the studies. Because of the large number of effect sizes and participants in this analysis, the finding of consistent deficits in face recognition is likely to generalize to the broader ASD population.

This finding is robust. We investigated the validity of the meta-analysis and its robustness to exaggerated effects from small studies that can have biased effects because of small sample sizes and biased sampling. This can be reflected in publication bias, the bias to publish studies on the basis of a single factor like statistical significance instead of multiple factors that influence effect size estimates, like statistical power, sample size, and overall quality of the study. Our analysis revealed that the full set of studies contained a small number of studies that were biased in this way. Importantly, even when we removed them from the analyses, the summary effects remained large, indicating that deficits in face recognition in ASD are robust across studies.

Does Methodological Variation Across Studies Influence Findings of Face Recognition Deficits in ASD?

In previous reviews of this literature, researchers hypothesized that variation in study findings may be due to heterogeneity in methodological approaches (Tang et al, 2015; Weigelt et al., 2012). Some of the proposed variations that likely affect the consistency of results across studies include the age, sex distribution, IQ, and sample sizes of the participant groups; the diagnostic criteria for the ASD group; the vast range of task paradigms; and overall design and reporting quality of the studies. We empirically investigated these hypotheses by evaluating the impact of these factors on the face identity recognition summary effect size.

Participant Age.—The majority of studies in this meta-analysis constrained the age range of the participants (e.g., children, adolescents, or adults), which allowed us to test whether differences in age group contributed to between-study heterogeneity and moderated the summary effect size. Importantly, there are only a small number of cross-sectional studies specifically evaluating age-related differences in face recognition abilities between ASD and TD individuals (Fedor et al., 2018; O'hearn et al., 2014; Scherf, Behrmann, Minshew, & Luna, 2008), and critically, no longitudinal studies, in this analysis. We identified 30 effect sizes representing contrasts between ASD and TD children (< 10 years), 47 effect sizes between ASD and TD adolescents (ages 10–18 years), and 38 effect sizes between ASD and TD adults (> 18 years). The meta-analysis revealed that face identity recognition deficits were not systematically impacted by the average age of the participants. Simply put, the magnitude of these deficits was similar across all age groups included in this meta-analysis. Although cross-sectional in nature, this suggests that the magnitude of face identity recognition deficits is persistent and consistent from childhood to middle adulthood.

Distribution of Sample Sex.—There is a growing notion that autistic symptomology may be manifest differently in men and women (e.g., Lai et al., 2017b), particularly in terms

of face processing (e.g. Whyte & Scherf, 2017). We tested this hypothesis by evaluating the extent to which face recognition deficits are impacted by the distribution of sex in the study samples (i.e., percentage of males in each study). The studies in this meta-analysis contained samples that ranged from 50% - 100% male; however, only two effect sizes were generated from samples that were 50% male. Given the strong male bias in the diagnosis of ASD (4:1), it is unsurprising that the majority of studies contained samples that have a majority percentage of males. The analyses revealed no systematic relation between the proportion of male participants in the study sample and the magnitude of face identity recognition deficits between ASD and TD groups. In other words, samples that contained a lower and higher proportion of males exhibited similar magnitude deficits in face identity recognition. However, it is important to keep in mind that the overwhelming majority of these studies included in the meta-analysis have a disproportionate number of males compared to females. Therefore, further research in which the distribution of males and females is consistently balanced is required to truly evaluate the possibility that the magnitude of face identity recognition deficits are modulated by sex.

Participant IQ.—In TD populations, IQ is uncorrelated with face identity recognition (Wilmer, Germine, & Nakayama, 2014; Zhu et al., 2010). There is no clear theoretical reason why that association would be different in ASD. Therefore, we would expect that participant IQ does not impact the magnitude of face identity recognition deficits between ASD and TD individuals. The current meta-analysis supported this notion. Specifically, we found that face identity recognition deficits in ASD were of similar magnitude regardless of the average IQ of the sample. Of note, although mean IQ scores ranged from 82 to 127, the majority of the samples reported average IQ scores between 100–110. Therefore, the summary effect size estimates we report is characteristic of individuals with IQ scores in the normal range and may not generalize to those with IQ scores outside that range.

Task Paradigm.—Face identity recognition was tested using a multitude of paradigms. For face recognition, these include tasks like “Old-New”, CFMT, and WMS. As part of the paradigm differences, studies also varied across task parameters such as the number of trials, distractors, time delay between study and response periods, and response time. In addition, the types of stimuli varied dramatically. For example, researchers used photographs that were either black/white or color, printed or digital images, spatially filtered (high and low pass), inverted, or partially occluded. Finally, it is likely that studies differ in a number of ways that are inaccessible to readers directly including, where and when the task was administered, the set of instructions prior to the start of the task, whether the task was administered in isolation or in the context of other tasks.

We coded the studies in this analysis on the basis of the type of task paradigm that was employed, which only captures some of this variability. The analysis revealed no evidence for an influence of task paradigm moderating the summary effect size of face identity recognition deficits between ASD and TD groups. This indicates that the magnitude of face identity recognition deficits in ASD persists regardless of the task used to assess face identity recognition behaviors. Importantly, this finding may be a function of the fact that we defined face identity recognition in a clear, operational way to determine the inclusion

criteria for the studies in this meta-analysis. This lends confidence to the external validity of these studies even though it has not been tested explicitly.

Design and Reporting Quality Index.—We tested the possibility that variations in the design and reporting quality across studies might influence the magnitude of the summary effect size on face identity recognition deficits in ASD. To do so, we derived an index of that included metrics of study design, reporting of participant demographics, and ASD diagnostic confirmation, such that higher scores indicated better design and reporting quality. In contrast to findings reported in a previous review (Tang et al., 2015), we did not observe an association between the extent of the design and reporting quality of the study the magnitude of the deficit in face identity recognition in ASD.

This lack of an association between DRQI scores and deficits may be related to the limited range in the scores across studies (see Supplementary Materials). Also, the use of these kinds of quality indices to weight studies in meta-analysis has actually been discouraged (Ahn & Becker, 2010; Herbison et al., 2006) because they only provide low dimensional information about study quality, which is determined by many dimensions.

Unexplained Heterogeneity Between Studies.—While we did find large, generalizable face identity recognition deficits in ASD, we also observed considerable heterogeneity among the reviewed studies. The difficult problem is trying to understand whether there is organized and meaningful variance in this heterogeneity that influences the magnitude of the summary effect size. One possibility is variance in task paradigms that were not evaluated here. For example, in unstandardized tasks like the “old/new” paradigm, studies routinely vary on the instructions for the task, number of trials, number of distractors, delay between study and test phases, and difficulty of the task. All of these (and other) subtle differences likely contribute to the between-study heterogeneity observed here. Moving forward, we recommend that the research community agree to using similar task paradigms and stimuli to minimize this between-study heterogeneity and compare results. As an example of this approach, we analyzed all the studies ($k = 20$) that used the CFMT paradigm, which is a highly regarded task of unfamiliar face identity recognition. Among this set of studies, the summary effect was large (Hedge’s $g = -0.84$, $se = .09$, 95% CI [−1.02, −0.67], $p < .0001$) and there was no significant heterogeneity among the observed effect sizes ($Q = 20.29$, $p = .38$; $I^2 = 18.55\%$).

Are Face Identity Processing Deficits in ASD Specific to Recognition?

Understanding the specific set of task demands that make processing face identity information especially difficult for individuals with ASD is important. Weigelt et al. (2012) reported that face identity recognition is deficient in ASD, but that perception of face identity information is intact. Furthermore, the researchers predicted that tasks with longer face memory demands will generate worse performance in ASD. If this is the case, there are important implications for developing targeted intervention programs aimed at improving face identity recognition abilities in ASD.

We empirically evaluated this hypothesis by computing and comparing the summary effect sizes from studies comparing face identity recognition ($k = 119$) and face identity

discrimination (k = 53) abilities between ASD and TD individuals. In contrast to predictions, our meta-analyses indicated that there are *comparable deficits* in both face identity recognition and discrimination abilities in ASD. As with face identity recognition deficits, the magnitude of the ASD deficit in face identity discrimination is nearly 1 SD below TD individuals. These findings suggest that the difficulty for ASD individuals has to do with processing information about individual face identity at both the perceptual and representational levels.

Finally, given the prominent deficits in face identity discrimination abilities in ASD, we evaluated whether methodological variation (as described above) influenced the estimated magnitude of this deficit. Comparable to our findings of the robustness of the estimated magnitude of deficits in face identity recognition, we did not observe any moderation of the summary effect size characterizing ASD deficits in face identity discrimination. This finding supports the notion that these perceptual and representational deficits in face identity processing are large and evident even in the face of so much methodological variation across studies.

Limitations

The meta-analytic approach is inherently limited by the available data from the existing literature. As a result, meta-analysis is influenced by the *file-drawer problem* (Rosenthal, 1979) and the *statistical significance filter*. The file-drawer problem refers to the bias introduced into the scientific literature resulting from the lower likelihood of nonsignificant and null findings to be published and available to the scientific community (i.e. “left in the file-drawer”). Recall that the goal of meta-analysis is to pool *all* data regardless of statistical significance in order to best represent the true population estimate. Without access to data “left in the file-drawer”, meta-analytic estimates can be biased.

Relatedly, the statistical significance filter also introduces a bias into the scientific literature. It reflects the phenomenon that statistically significant findings tend to overestimate the magnitude of an effect (Type M Errors; Gelman & Carlin, 2014). This is because statistically significant (i.e., $p < .05$) findings are much more likely to be published. Such results are also likely to be large (±1.96 standard deviations) in order to reach the significance threshold ($p < .05$). As a result, it is possible that only large effects get published, leading to Type M errors. This is especially critical when studies are underpowered because the effects have to be substantially large in order to reach statistical significance with a small sample size. This is why it is important to evaluate publication bias, which is a general term reflecting the tendency of publishers to only publish “positive” or “novel” findings. This is also why investigation of funnel plots has become a standard method for evaluating potential sensitivity to small-study effects. For example, it is almost always the case that wildly large effects (e.g., Cohen’s $d > 5$) are estimated from small samples with a large sampling error, but these studies get published because they report statistically significant results (i.e., pass the statistical significance filter).

This problem is further complicated by the fact that there are limited statistical methods for evaluating publication bias with multilevel data (Nakagawa & Santos, 2012). Nevertheless, we employed a solution to evaluate funnel plot asymmetry and found only a small number

of effect sizes that were prone to small-study effects, like publication bias, and were likely influencing the results in an extreme way. Importantly, upon removal, the summary effect sizes for face identity recognition and discrimination were still large, and only minimally reduced, indicating that the findings from this meta-analysis are robust to these potential sources of bias.

Finally, these findings are limited by the research questions that have been addressed and experimental paradigms that have been used in the current literature. For example, all of the studies we included in the meta-analysis investigated face identity recognition or discrimination of *unfamiliar* faces. Therefore, the findings from this work only generalize to characterize unfamiliar face identity processing deficits in ASD. The literature evaluating the face identity recognition and discrimination of *familiar* faces in ASD is miniscule in comparison, but does indicate similar deficits (e.g., Boucher et al., 1998; Pierce & Redcay, 2008; Sterling et al., 2008; Wilson et al., 2007). Future work employing the meta-analytic approach to the findings of these studies will be important given existing work suggesting that the processes and neural systems supporting *familiar* face identity processing may be overlapping, but different, from those for unfamiliar face identity processing (Ramon & Gobbini, 2018; but see also Abudarham et al., 2019).

Implications of Deficient Face Identity Processing for Autism

These findings indicate that individuals with ASD have difficulties processing information about face identity. These deficits are large and evident in tests of both face identity recognition and discrimination. Face identity processing deficits in ASD are present in males and females, across a wide range of ages and IQ (within normal range). Although these deficits appear to affect most people with ASD, they are not considered core *clinical symptoms* of ASD. In other words, difficulties perceiving and recognizing individual faces are not part of the clinical criteria for diagnosing ASD. However, deficient face identity processing may underlie many of the atypical behaviors that represent core ASD symptoms in the social communication domain. For example, difficulty keeping track of face identity might contribute to abnormal social approach behavior and inappropriate social interactions. In turn, these atypical social behaviors likely make it more difficult for individuals with ASD to develop and maintain social relationships and to adjust their behavior in socially appropriate ways.

Consistent with this notion, there is some evidence that the magnitude of deficits in face identity recognition abilities is related to social-communicative difficulties in ASD. For example, the severity of face recognition deficits in ASD has been associated with reduced social engagement and cooperative play (Corbett, Newsom, Key, Qualls, & Edmiston, 2014) and lower social functioning (Neuhaus et al., 2016; Schultz, 2005; Webb, Neuhaus, & Faja, 2017); these deficits are also associated with increased social withdrawal (Anderson, Maye, & Lord, 2011). Perhaps most convincingly, a 7-year longitudinal study of 87 children (6–12-year-olds) found that face recognition behavior in childhood predicted the severity of ADOS scores later in adolescence (12–19-year-olds; Eussen et al., 2015). These findings indicate that face recognition deficits have clinical relevance for ASD, particularly in terms of social communicative behavior.

Conclusion

By aggregating data from 40 years of research, this quantitative meta-analysis estimates that face identity processing deficits in autism are consistent and large. Specifically, individuals with ASD perform approximately 1 SD worse than their typically developing peers in tasks of face identity recognition and discrimination. We also report that the magnitude of these effect sizes is consistent across variation in age, sex, IQ, task paradigm, and design and reporting quality. Although atypical face identity processing is not currently part of the diagnostic criteria for autism, this work indicates that these deficits are pervasive and large. This information will be useful for those developing targeted interventions for individuals with ASD and those investigating underlying mechanisms that contribute to this difficulty in processing face identity information at both the perceptual and representational levels in ASD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by grants from the National Institute of Mental Health R61/33 MH11-624 (KSS).

References

*Indicates that these studies were included in the meta-analysis

Abudarham N, Shkiller L, Yovel G. (2019). Critical features for face recognition Cognition, 182, 73–83. 10.1016/j.cognition.2018.09.002 [PubMed: 30218914]

Ahn S, & Becker BJ (2010). Incorporating Quality Scores in Meta-Analysis. Journal of Educational and Behavioral Statistics, 36(5), 555–585. doi: 10.3102/1076998610393968

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th Ed.). Washington, DC: Author.

Anderson DK, Maye MP, & Lord C. (2011). Changes in Maladaptive Behaviors From Mid childhood to Young Adulthood in Autism Spectrum Disorder. American Journal on Intellectual and Developmental Disabilities, 116(5), 381–397. doi: 10.1352/1944-7558-116.5.381 [PubMed: 21905806]

*Annaz D, Karmiloff-Smith A, Johnson MH, & Thomas MS (2009). A cross-syndrome study of the development of holistic face recognition in children with autism, Down syndrome, and Williams syndrome. Journal of Experimental Child Psychology, 102(4), 456–486. doi: 10.1016/j.jecp.2008.11.005 [PubMed: 19193384]

*Arkush L, Smith-Collins AP, Fiorentini C, & Skuse DH (2013). Recognition of face and non-face stimuli in autistic spectrum disorder. Autism Research : Official Journal of the International Society for Autism Research, 6(6), 550–560. doi: 10.1002/aur.1318 [PubMed: 23894016]

Assink M, & Wibbelink CJ (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. The Quantitative Methods for Psychology, 12(3), 154–174. doi: 10.20982/tqmp.12.3.p154

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, & Clubley E. (2001). The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31(1), 5–17. doi: 10.1023/a:1005653411471 [PubMed: 11439754]

*Barron-Linnankoski S, Reinvall O, Lahervuori A, Voutilainen A, Lahti-Nuuttila P, & Korkman M. (2014). Neurocognitive performance of children with higher functioning Autism Spectrum disorders

on the NEPSY-II. Child Neuropsychology, 21(1), 55–77. doi: 10.1080/09297049.2013.873781 [PubMed: 24397431]

Benton AL, Sivan AB, Hamsher K, Varney NR, & Spreen O. (1983). Contribution to neuropsychological assessment. NY: Oxford University Press; 1983.

*Blair R, Frith U, Smith N, Abell F, & Cipolotti L. (2002). Fractionation of visual memory: agency detection and its impairment in autism. Neuropsychologia, 40(1), 108–118. 10.1016/s0028-3932(01)00069-0 [PubMed: 11595266]

*Bookheimer S, Wang A, Scott A, Sigman M, Dapretto M. (2008). Frontal contributions to face processing differences in autism: Evidence from fMRI of inverted face processing. Journal of the International Neuropsychological Society 14(6), 922–932. 10.1017/s135561770808140x [PubMed: 18954473]

*Boucher J, & Lewis V. (1992). Unfamiliar face recognition in relatively able autistic children. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 33(5), 843–859. doi: 10.1111/j.1469-7610.1992.tb01960.x

Boucher J, Lewis V, Collis G. (1998). Familiar face and voice matching and recognition in children with autism. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 39(2), 171—181. 10.1111/1469-7610.00311

Bruce V, Bindemann M, & Lander K. (2018). Individual differences in face perception and person recognition. Cognitive Research: Principles and Implications, 3(1), 18. doi: 10.1186/s41235-018-0109-4 [PubMed: 30009248]

Burton A, Wilson S, Cowan M, Bruce V. (1999). Face recognition in poor-quality video: evidence from security surveillance. Psychological Science, 10(3), 243–248. 10.1111/1467-9280.00144

*Campbell R, Lawrence K, William M, Cheytna M, Jeyakuma L, & Skuse D. (2006). Meanings in motion and faces: developmental associations between the processing of intention from geometrical animations and gaze detection accuracy. Development and Psychopathology, 18(1), 99–118. doi: 10.1017/s0954579406060068 [PubMed: 16478554]

*Celani G, Battacchi M, & Arcidiacono L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. Journal of Autism and Developmental Disorders, 29(1), 57–66. doi: 10.1023/a:1025970600181 [PubMed: 10097995]

Cheung M. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. Psychological Methods, 19(2), 211. doi: 10.1037/a0032968 [PubMed: 23834422]

Cheung M. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. Neuropsychology Review, 29, 387–396. 10.1007/s11065-019-09415-6 [PubMed: 31446547]

*Chien H-L, Wang L-H, Chen C-C, Chen T-Y, & Chen H-S (2014). Autistic children do not exhibit an own-race advantage as compared to typically developing children. Research in Autism Spectrum Disorders, 8(11), 1544–1551. doi: 10.1016/j.rasd.2014.08.005

Chita-Tegmark M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. Research in Developmental Disabilities, 48, 79–93. doi: 10.1016/j.ridd.2015.10.011 [PubMed: 26547134]

*Conturo TE, Williams DL, Smith CD, Gultepe E, Akbudak E, & Minshew NJ (2008). Neuronal fiber pathway abnormalities in autism: An initial MRI diffusion tensor tracking study of hippocampo-fusiform and amygdalo-fusiform pathways. Journal of the International Neuropsychological Society, 14(6), 933–946. doi: 10.1017/s1355617708081381 [PubMed: 18954474]

*Corbett BA, Newsom C, Key AP, Qualls LR, & Edmiston KE (2014). Examining the relationship between face processing and social interaction behavior in children with and without autism spectrum disorder. Journal of Neurodevelopmental Disorders, 6(1), 35. doi: 10.1186/1866-1955-6-35 [PubMed: 25180050]

Cumming G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Routledge.

*Dalton KM, Nacewicz BM, Alexander AL, & Davidson RJ (2007). Gaze-fixation, brain activation, and amygdala volume in unaffected siblings of individuals with autism. Biological Psychiatry, 61(4), 512–520. doi: 10.1016/j.biopsych.2006.05.019 [PubMed: 17069771]

*Davies S, Bishop D, Manstead A, Tantam D. (1994). Face perception in children with autism and Asperger's syndrome. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 35(6), 1033–1057. 10.1111/j.1469-7610.1994.tb01808.x

Dawson G, Carver L, Meltzoff A, Panagiotides H, McPartland J, Webb S. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. Child development, 73(3), 700—717. 10.1111/1467-8624.00433 [PubMed: 12038546]

*de Gelder B, Vroomen J, & van der Heide L. (1991). Face recognition and lip-reading in autism. European Journal of Cognitive Psychology, 3(1), 69–86. doi: 10.1080/09541449108406220

*Deruelle C, Rondan C, Gepner B, Tardif C. (2004). Spatial frequency and face processing in children with autism and Asperger syndrome. Journal of Autism and Developmental Disorders, 34(2), 199–210. 10.1023/b:jadd.0000022610.09668.4c [PubMed: 15162938]

*Dimitriou D, Leonard H, Karmiloff-Smith A, Johnson M, & Thomas C. (2015). Atypical development of configural face recognition in children with autism, Down syndrome and Williams syndrome. Journal of Intellectual Disability Research, 59(5), 422–438. doi: 10.1111/jir.12141 [PubMed: 25059077]

Duchaine B, & Nakayama K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia, 44(4), 576–585. doi: 10.1016/j.neuropsychologia.2005.07.001 [PubMed: 16169565]

Egger M, Smith G, Schneider M, Minder C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ, 315(7109), 629–634. 10.1136/bmj.315.7109.629 [PubMed: 9310563]

Elbich DB, & Scherf SK (2017). Beyond the FFA: Brain-behavior correspondences in face recognition abilities. NeuroImage, 147, 409–422. doi: 10.1016/j.neuroimage.2016.12.042 [PubMed: 27993674]

Eussen ML, Louwerse A, Herba CM, Gool AR, Verheij F, Verhulst FC, & Greaves-Lord K. (2015). Childhood facial recognition predicts adolescent symptom severity in autism spectrum disorder. Autism Research, 8(3), 261–271. doi: 10.1002/aur.1443 [PubMed: 25631781]

*Ewbank MP, Pell PJ, Powell TE, von dem Hagen EA, Baron-Cohen S, & Calder AJ (2017). Repetition suppression and memory for faces is reduced in adults with autism spectrum conditions. Cerebral Cortex, 27(1), 92–103. doi: 10.1093/cercor/bhw373 [PubMed: 27909005]

*Ewing L, Pellicano E, King H, Lennuyeux-Comnene L, Farran EK, Karmiloff-Smith A, & Smith ML (2018). Atypical information-use in children with autism spectrum disorder during judgments of child and adult face identity. Developmental Neuropsychology, 43(4), 370–384. doi: 10.1080/87565641.2018.1449846 [PubMed: 29558171]

*Ewing L, Pellicano E, & Rhodes G. (2013a). Atypical updating of face representations with experience in children with autism. Developmental Science, 16(1), 116–123. doi: 10.1111/desc.12007 [PubMed: 23278933]

*Ewing L, Pellicano E, & Rhodes G. (2013b). Reevaluating the selectivity of face-processing difficulties in children and adolescents with autism. Journal of Experimental Child Psychology, 115(2), 342–355. doi: 10.1016/j.jecp.2013.01.009 [PubMed: 23563163]

*Ewing L, Pellicano E, & Rhodes G. (2013c). Using effort to measure reward value of faces in children with autism. PloS One, 8(11), e79493. doi: 10.1371/journal.pone.0079493

*Faja S, Webb S, Merkle K, Aylward E, & Dawson G. (2009). Brief report: face configuration accuracy and processing speed among adults with high-functioning autism spectrum disorders. Journal of Autism and Developmental Disorders, 39(3), 532–538. doi: 10.1007/s10803-008-0635-x [PubMed: 18751880]

*Fedor J, Lynn A, Foran W, DiCicco-Bloom J, Luna B, & O'Hearn K. (2018). Patterns of fixation during face recognition: Differences in autism across age. Autism, 22(7), 866–880. doi: 10.1177/1362361317714989 [PubMed: 28782371]

*Fein D, Barton M, Eigsti I, Kelley E, Naigles L, Schultz RT, … Tyson K. (2013). Optimal outcome in individuals with a history of autism. Journal of Child Psychology and Psychiatry, 54(2), 195–205. doi: 10.1111/jcpp.12037 [PubMed: 23320807]

Frazier TW, Strauss M, Klingemier EW, Zetzer EE, Hardan AY, Eng C, & Youngstrom EA (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. Journal of the American Academy of Child & Adolescent Psychiatry, 56(7), 546–555. doi: 10.1016/j.jaac.2017.05.005 [PubMed: 28647006]

Fysh MC (2018). Individual differences in the detection, matching and memory of faces. Cognitive Research: Principles and Implications, 3(1), 20. doi: 10.1186/s41235-018-0111-x

Gelman A, & Carlin J. (2014). Beyond power calculations. Perspectives on Psychological Science, 9(6), 641–651. doi: 10.1177/1745691614551642 [PubMed: 26186114]

*Gepner B, de Gelder B, & de Schonen S. (1996). Face processing in autistics: Evidence for a generalised deficit? Child Neuropsychology, 2(2), 123–139. doi: 10.1080/09297049608401357

Germine LT, Duchaine B, & Nakayama K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. Cognition, 118(2), 201–210. doi: 10.1016/j.cognition.2010.11.002 [PubMed: 21130422]

*Greimel E, Schulte-Ruther M, Kamp-Becker I, Remschmidt H, Herpertz-Dahlmann B, & Konrad K. (2014). Impairment in face processing in autism spectrum disorder: a developmental perspective. Journal of Neural Transmission (Vienna, Austria : 1996), 121(9), 1171–1181. doi: 10.1007/s00702-014-1206-2

Griffin JW (2020). metapoweR: an R package for computing meta-analytic statistical power. R package version 0.2.0. https://CRAN.R-project.org/package=metapower.

Halliday DW, MacDonald SW, Scherf KS, & Tanaka JW (2014). A reciprocal model of face recognition and autistic traits: evidence from an individual differences perspective. PLoS ONE, 9(5), e94013. doi: 10.1371/journal.pone.0094013

*Hauck M, Fein D, Maltby N, Waterhouse L, & Feinstein C. (1998). Memory for faces in children with autism. Child Neuropsychology (Neuropsychology, Development and Cognition: Section C), 4(3), 187–198. doi: 10.1076/chin.4.3.187.3174

Hedges LV (1981). Distribution theory for glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6(2), 107–128. doi: 10.3102/10769986006002107

*Hedley D, Brewer N, & Young R. (2011). Face recognition performance of individuals with Asperger syndrome on the Cambridge Face Memory Test. Autism Research, 4(6), 449–455. doi: 10.1002/aur.214 [PubMed: 22162360]

*Hedley D, Brewer N, & Young R. (2015). The effect of inversion on face recognition in adults with autism spectrum disorder. Journal of Autism and Developmental Disorders, 45(5), 1368–1379. doi: 10.1007/s10803-014-2297-1 [PubMed: 25358250]

*Hedley D, Young R, & Brewer N. (2012). Using eye movements as an index of implicit face recognition in autism spectrum disorder. Autism Research, 5(5), 363–379. doi: 10.1002/aur.1246 [PubMed: 22865711]

Herbison P, Hay-Smith J, & Gillespie WJ (2006). Adjustment of meta-analyses on the basis of quality scores should be abandoned. Journal of Clinical Epidemiology, 59(12), 1249.e1–1249.e11. doi: 10.1016/j.jclinepi.2006.03.008

*Herrington J, Miller J, Pandey J, Schultz R. (2016). Anxiety and social deficits have distinct relationships with amygdala function in autism spectrum disorder. Social Cognitive and Affective Neuroscience, 11(6), 907—914. 10.1093/scan/nsw015 [PubMed: 26865425]

*Herrington JD, Riley ME, Grupe DW, & Schultz RT (2015). Successful face recognition is associated with increased prefrontal cortex activation in autism spectrum disorder. Journal of Autism and Developmental Disorders, 45(4), 902–910. doi: 10.1007/s10803-014-2233-4 [PubMed: 25234479]

*Hooper SR, Poon KK, Marcus L, & Fine C. (2006). Neuropsychological characteristics of school-age children with high-functioning autism: Performance on the NEPSY. Child Neuropsychology, 12(4–5), 299–305. doi: 10.1080/09297040600737984 [PubMed: 16911974]

*Howard M, Cowell P, Boucher J, Broks P, Mayes A, Farrant A, & Roberts N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. NeuroReport, 11(13), 2931–2935. doi: 10.1097/00001756-200009110-00020 [PubMed: 11006968]

*Humphreys K, Minshew N, Leonard G, & Behrmann M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. Neuropsychologia, 45(4), 685–695. doi: 10.1016/j.neuropsychologia.2006.08.003 [PubMed: 17010395]

Hunter JE, & Schmidt FL (1990). Methods of meta-analysis: Correcting error and bias in research findings. Thousand Oaks, CA, US: Sage Publications, Inc.

*Ipser A, Ring M, Murphy J, Gaigg SB, & Cook R. (2016). Similar exemplar pooling processes underlie the learning of facial identity and handwriting style: Evidence from typical observers and individuals with autism. Neuropsychologia, 85, 169–176. doi: 10.1016/j.neuropsychologia.2016.03.017 [PubMed: 27001029]

Jenkins R, Dowsett A, & Burton A. (2018). How many faces do people know? Proceedings of the Royal Society B, 285(1888), 20181319. doi: 10.1098/rspb.2018.1319

*Jiang YV, Palm BE, DeBolt MC, & Goh Y. (2015). High-precision visual long-term memory in children with high-functioning autism. Journal of Abnormal Psychology, 124(2), 447–456. doi: 10.1037/abn0000022 [PubMed: 25436998]

*Jones RR, Blades M, Coleman M, & Pascalis O. (2013). Learning new faces in typical and atypical populations of children. Scandinavian Journal of Psychology, 54(1), 10–13. doi: 10.1111/sjop.12006 [PubMed: 23320881]

*Joseph RM, Ehrman K, McNally R, & Keehn B. (2008). Affective response to eye contact and face recognition ability in children with ASD. Journal of the International Neuropsychological Society, 14(6), 947–955. doi: 10.1017/s1355617708081344 [PubMed: 18954475]

*Joseph R. & Tanaka J. (2003). Holistic and part-based face recognition in children with autism. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 44(4), 529–542. 10.1111/1469-7610.00142

*Jung CE, Strother L, Feil-Seifer DJ, & Hutsler JJ (2016). Atypical asymmetry for processing human and robot faces in autism revealed by fNIRS. PLOS ONE, 11(7), e0158804. doi: 10.1371/journal.pone.0158804

*Key AP, & Corbett BA (2014). ERP responses to face repetition during passive viewing: a nonverbal measure of social motivation in children with autism and typical development. Developmental Neuropsychology, 39(6), 474–495. doi: 10.1080/87565641.2014.940620 [PubMed: 25144259]

*Kirchner JC, Hatri A, Heekeren HR, & Dziobek I. (2011). Autistic symptomatology, face processing abilities, and eye fixation patterns. Journal of Autism and Developmental Disorders, 41(2), 158–167. doi: 10.1007/s10803-010-1032-9 [PubMed: 20503068]

*Kleinhans NM, Johnson CL, Richards T, Mahurin R, Greenson J, Dawson G, & Aylward E. (2009). Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. American Journal of Psychiatry, 166(4), 467–475. doi: 10.1176/appi.ajp.2008.07101681

Kmet L, Lee R, & Cook L. (2004). Standard quality assessment criteria for evaluating primary research papers from a variety of fields (HTA Initiative #13). Edmonton, AB: Alberta Heritage Foundation for Medical Research.

Korkman M. Kirk U, & Kemp S. (2007). NEPSY (2nd ed.) San Antonio, TX: Harcourt Assessment

*Koshino H, Kana RK, Keller TA, Cherkassky VL, Minshew NJ, & Just M. (2008). fMRI investigation of working memory for faces in autism: visual coding and underconnectivity with frontal areas. Cerebral Cortex, 18(2), 289–300. doi: 10.1093/cercor/bhm054 [PubMed: 17517680]

*Krysko K, Rutherford M. (2009). A threat-detection advantage in those with autism spectrum disorders. Brain and Cognition, 69(3), 472–480. 10.1016/j.bandc.2008.10.002 [PubMed: 19036491]

*Kuusikko-Gauffin S, Eira J-V, Alice C, Rachel P-W, Katja J, Marja-Leena M, … Irma M. (2011). Face memory and object recognition in children with high-functioning autism or Asperger syndrome and in their parents. Research in Autism Spectrum Disorders, 5(1), 622–628. doi: 10.1016/j.rasd.2010.07.007

*Kylliäinen A, Braeutigam S, Hietanen JK, Swithenby SJ, & Bailey AJ (2006). Face- and gaze-sensitive neural responses in children with autism: a magnetoencephalographic study. European Journal of Neuroscience, 24(9), 2679–2690. doi: 10.1111/j.1460-9568.2006.05132.x

Lai C, Lau Z, Lui SS, Lok E, Tam V, Chan Q, … Cheung EF (2017a). Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. Autism Research, 10(5), 911–939. doi: 10.1002/aur.1723 [PubMed: 27874266]

Lai M-C, Lombardo MV, Ruigrok A, Chakrabarti B, Auyeung B, Szatmari P, … Baron-Cohen S. (2017b). Quantifying and exploring camouflaging in men and women with autism. Autism : The International Journal of Research and Practice, 21(6), 690–702. doi: 10.1177/1362361316671012

*Lajiness-O'Neill RR, Beaulieu I, Titus JB, Asamoah A, Bigler ED, Bawle EV, & Pollack R. (2005). Memory and learning in children with 22q11.2 deletion syndrome: evidence for ventral and dorsal stream disruption? Child Neuropsychology, 11(1), 55–71. doi: 10.1080/09297040590911202 [PubMed: 15823983]

Langdell T. (1978). Recognition of faces: an approach to the study of autism. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 19(3), 255–268. doi: 10.1111/j.1469-7610.1978.tb00468.x

*Leonard HC, Annaz D, Karmiloff-Smith A, & Johnson MH (2011). Developing spatial frequency biases for face recognition in autism and Williams syndrome. Journal of Autism and Developmental Disorders, 41(7), 968–973. doi: 10.1007/s10803-010-1115-7 [PubMed: 20945155]

*Li T, Wang X, Pan J, Feng S, Gong M, Wu Y, … Yi L. (2017). Reward learning modulates the attentional processing of faces in children with and without autism spectrum disorder. Autism Research, 10(11), 1797–1807. doi: 10.1002/aur.1823 [PubMed: 28608397]

*López B, Donnelly N, Hadwin J, Leekam S. (2004). Face processing in high-functioning adolescents with autism: Evidence for weak central coherence. Visual Cognition, 11(6), 673–688. 10.1080/13506280344000437

*López B, Leekam SR, & Arts GR (2008). How central is central coherence? Preliminary evidence on the link between conceptual and perceptual processing in children with autism. Autism, 12(2), 159–171. doi: 10.1177/1362361307086662 [PubMed: 18308765]

Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL,et al. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders, 30, 205–223. [PubMed: 11055457]

Lord C, Rutter M, Le Couteur A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders, 24, 659–685. [PubMed: 7814313]

*Lynn AC, Padmanabhan A, Simmonds D, Foran W, Hallquist MN, Luna B, & O'Hearn K. (2018). Functional connectivity differences in autism during face and car recognition: underconnectivity and atypical age-related changes. Developmental Science, 21(1). doi: 10.1111/desc.12508

*McPartland JC, Webb SJ, Keehn B, & Dawson G. (2011b). Patterns of visual attention to faces and objects in autism spectrum disorder. Journal of Autism and Developmental Disorders, 41(2), 148–157. doi: 10.1007/s10803-010-1033-8 [PubMed: 20499148]

*McPartland JC, Dawson G, Webb SJ, Panagiotides H, & Carver LJ (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 45(7), 1235–1245. doi: 10.1111/j.1469-7610.2004.00318.x

*McPartland JC, Wu J, Bailey CA, Mayes LC, Schultz RT, & Klin A. (2011a). Atypical neural specialization for social percepts in autism spectrum disorder. Social Neuroscience, 6(5–6), 436–451. doi: 10.1080/17470919.2011.586880 [PubMed: 21777159]

Moher D, Liberati A, Tetzlaff J, & Altman DG (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097. 10.1371/journal.pmed.1000097

Nakagawa S, & Santos ES (2012). Methodological issues and advances in biological meta-analysis. Evolutionary Ecology, 26(5), 1253–1274. doi: 10.1007/s10682-012-9555-5

*Narzisi A, Muratori F, Calderoni S, Fabbro F, & Urgesi C. (2013). Neuropsychological profile in high functioning autism spectrum disorders. Journal of Autism and Developmental Disorders, 43(8), 1895–1909. doi: 10.1007/s10803-012-1736-0 [PubMed: 23224514]

*Neil L, Cappagli G, Karaminis T, Jenkins R, Pellicano E. (2016). Recognizing the same face in different contexts: Testing within-person face recognition in typical development and in autism. Journal of Experimental Child Psychology, 143, 139—153. 10.1016/j.jecp.2015.09.029 [PubMed: 26615971]

Neuhaus E, Kresse A, Faja S, Bernier RA, & Webb S. (2016). Face processing among twins with and without autism: social correlates and twin concordance. Social Cognitive and Affective Neuroscience, 11(1), 44–54. doi: 10.1093/scan/nsv085 [PubMed: 26137974]

*O'Brien J, Spencer J, Girges C, Johnston A, & Hill H. (2014). Impaired perception of facial motion in autism spectrum disorder. PloS One, 9(7), e102173. doi: 10.1371/journal.pone.0102173

*Oerlemans AM, Droste K, van Steijn DJ, de Sonneville LM, Buitelaar JK, & Rommelse NN (2013). Co-segregation of social cognition, executive function and local processing style in children with ASD, their siblings and normal controls. Journal of Autism and Developmental Disorders, 43(12), 2764–2778. doi: 10.1007/s10803-013-1807-x [PubMed: 23532348]

*O'Hearn K, Schroer E, Minshew N, & Luna B. (2010). Lack of developmental improvement on a face memory task during adolescence in autism. Neuropsychologia, 48(13), 3955–3960. doi: 10.1016/j.neuropsychologia.2010.08.024 [PubMed: 20813119]

*O'Hearn K, Tanaka J, Lynn A, Fedor J, Minshew N, & Luna B. (2014). Developmental plateau in visual object processing from adolescence to adulthood in autism. Brain and Cognition, 90, 124–134. doi: 10.1016/j.bandc.2014.06.004 [PubMed: 25019999]

O'Toole A, Phillips P, Weimer S, Roark D, Ayyad J, Barwick R, Dunlop J. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. Vision Research, 51(1), 74–83. 10.1016/j.visres.2010.09.035 [PubMed: 20969886]

*Ozonoff S, Pennington B, Rogers S. (1990). Are there emotion perception deficits in young autistic children? Journal of Child Psychology and Psychiatry, 31(3), 343–361. 10.1111/j.1469-7610.1990.tb01574.x [PubMed: 2318918]

*Philip RCM, Whalley HC, Stanfield AC, Sprengelmeyer R, Santos IM, Young AW, … Hall J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. Psychological Medicine, 40(11), 1919–1929. doi: 10.1017/s0033291709992364 [PubMed: 20102666]

Picci G, & Scherf SK (2014). A two-hit model of autism. Clinical Psychological Science, 3(3), 349–371. doi: 10.1177/2167702614540646 [PubMed: 26609500]

*Pierce K, & Redcay E. (2008). Fusiform function in children with an autism spectrum disorder is a matter of "who". Biological Psychiatry, 64(7), 552–560. doi: 10.1016/j.biopsych.2008.05.013 [PubMed: 18621359]

*Planche P, & Lemonnier E. (2012). Children with high-functioning autism and Asperger's syndrome: Can we differentiate their cognitive profiles? Research in Autism Spectrum Disorders, 6(2), 939–948. doi: 10.1016/j.rasd.2011.12.009

Ramon M, Gobbini M. (2017). Familiarity matters: A review on prioritized processing of personally familiar faces. Visual Cognition, 26(3), 1–17. 10.1080/13506285.2017.1405134

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

*Reinvall O, Voutilainen A, Kujala T, & Korkman M. (2013). Neurocognitive functioning in adolescents with autism spectrum disorder. Journal of Autism and Developmental Disorders, 43(6), 1367–1379. doi: 10.1007/s10803-012-1692-8 [PubMed: 23104618]

*Rhodes G, Ewing L, Jeffery L, Avard E, & Taylor L. (2014). Reduced adaptability, but no fundamental disruption, of norm-based face-coding mechanisms in cognitively able children and adolescents with autism. Neuropsychologia, 62, 262–268. doi: 10.1016/j.neuropsychologia.2014.07.030 [PubMed: 25090925]

*Rhodes G, Neumann MF, Ewing L, & Palermo R. (2015). Reduced set averaging of face identity in children and adolescents with autism. Quarterly Journal of Experimental Psychology (2006), 68(7), 1391–1403. doi: 10.1080/17470218.2014.981554 [PubMed: 25510166]

*Riby D, Doherty-Sneddon G, Bruce V. (2008). Exploring face perception in disorders of development: Evidence from Williams syndrome and autism. Journal of Neuropsychology, 2(1), 47–64. 10.1348/174866407x255690 [PubMed: 19334304]

*Riby D, Doherty-Sneddon G, Bruce V. (2009). The eyes or the mouth? Feature salience and unfamiliar face processing in Williams syndrome and autism. The Quarterly Journal of Experimental Psychology, 62(1), 189–203. 10.1080/17470210701855629 [PubMed: 18609381]

Rice A, Phillips PJ, Natu V, An X, & O'Toole AJ (2013). Unaware person recognition from the body when face identification fails. Psychological Science, 24(11), 2235–2243. 10.1177/0956797613492986 [PubMed: 24068115]

*Rigby S, Stoesz B, Jakobson L. (2018). Empathy and face processing in adults with and without autism spectrum disorder. Autism Research, 11(6), 942–955. 10.1002/aur.1948 [PubMed: 29637718]

*Robel L, Ennouri K, Piana H, Vaivre-Douret L, Perier A, Flament MF, & Mouren-Simeoni MC (2004). Discrimination of face identities and expressions in children with autism: same or different? European Child & Adolescent Psychiatry, 13(4), 227–233. doi: 10.1007/s00787-004-0409-8 [PubMed: 15365893]

*Rondan C, Gepner B, Deruelle C. (2003). Inner and outer face perception in children with autism. Child Neuropsychology 9(4), 289–297. 10.1076/chin.9.4.289.23516 [PubMed: 14972707]

*Rose FE, Lincoln AJ, Lai Z, Ene M, Searcy YM, & Bellugi U. (2007). Orientation and affective expression effects on face recognition in Williams syndrome and autism. Journal of Autism and Developmental Disorders, 37(3), 513–522. doi: 10.1007/s10803-006-0200-4 [PubMed: 16906460]

Rosenthal R. (1979). The file drawer problem and tolerance for null results. Psychological Bulletin, 86(3), 638. doi: 10.1037/0033-2909.86.3.638

*Rosset D, Santos A, Fonseca D, Poinso F, O'Connor K, Deruelle C. (2010). Do children perceive features of real and cartoon faces in the same way? Evidence from typical development and autism. Journal of Clinical and Experimental Neuropsychology, 32(2), 212–218. 10.1080/13803390902971123 [PubMed: 19562609]

Rutter M, Bailey A, & Lord C. (2003). The social communication questionnaire: manual. Los Angeles, CA: Western Psychological Services.

*Sachse M, Schlitt S, Hainz D, Ciaramidaro A, Walter H, Poustka F, … Freitag CM (2014). Facial emotion recognition in paranoid schizophrenia and autism spectrum disorder. Schizophrenia Research, 159(2–3), 509–514. doi: 10.1016/j.schres.2014.08.030 [PubMed: 25278104]

*Schelinski S, Riedel P, & von Kriegstein K. (2014). Visual abilities are important for auditory-only speech recognition: Evidence from autism spectrum disorder. Neuropsychologia, 65, 1–11. doi: 10.1016/j.neuropsychologia.2014.09.031 [PubMed: 25283605]

*Schelinski S, Roswandowitz C, & von Kriegstein K. (2017). Voice identity processing in autism spectrum disorder. Autism Research, 10(1), 155–168. doi: 10.1002/aur.1639 [PubMed: 27404447]

*Scherf KS, Behrmann M, Minshew N, & Luna B. (2008). Atypical development of face and greeble recognition in autism. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 49(8), 838–847. doi: 10.1111/j.1469-7610.2008.01903.x

*Scherf KS, Elbich D, Minshew N, & Behrmann M. (2015). Individual differences in symptom severity and behavior predict neural activation during face processing in adolescents with autism. NeuroImage: Clinical, 7, 53–67. doi: 10.1016/j.nicl.2014.11.003

Schultz RT (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. International Journal of Developmental Neuroscience, 23(2–3), 125–141. doi: 10.1016/j.ijdevneu.2004.12.012 [PubMed: 15749240]

*Schultz R, Gauthier I, Klin A, Fulbright R, Anderson A, Volkmar F, Skudlarski P, Lacadie C, Cohen D, & Gore J. (2000). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome. Archives of General Psychiatry, 57(4), 331–340. 10.1001/archpsyc.57.4.331 [PubMed: 10768694]

*Serra M, Althaus M, Sonneville L, Stant A, Jackson A, Minderaa R. (2003). Face recognition in children with a pervasive developmental disorder not otherwise specified. Journal of Autism and Developmental Disorders, 33(3), 303–317. 10.1023/a:1024458618172 [PubMed: 12908833]

*Snow J, Ingeholm JE, Levy IF, Caravella RA, Case LK, Wallace GL, & Martin A. (2011). Impaired visual scanning and memory for faces in high-functioning autism spectrum disorders: It's not just the eyes. Journal of the International Neuropsychological Society, 17(6), 1021–1029. doi: 10.1017/s1355617711000981 [PubMed: 21892988]

*Southwick JS, Bigler ED, Froehlich A, DuBray MB, Alexander AL, … Lainhart JE (2011). Memory functioning in children and adolescents with autism. Neuropsychology, 25(6), 702–710. doi: 10.1037/a0024935 [PubMed: 21843004]

*Sterling L, Dawson G, Webb S, Murias M, Munson J, Panagiotides H, & Aylward E. (2008). The role of face familiarity in eye tracking of faces by individuals with autism spectrum disorders. Journal of Autism and Developmental Disorders, 38(9), 1666–1675. doi: 10.1007/s10803-008-0550-1 [PubMed: 18306030]

Tang J, Falkmer M, Horlin C, Tan T, Vaz S, & Falkmer T. (2015). Face recognition and visual search strategies in autism spectrum disorders: amending and extending a recent review by Weigelt et al. PLOS ONE, 10(8), e0134439. doi: 10.1371/journal.pone.0134439

*Tehrani-Doost M, Salmanian M, Ghanbari-Motlagh M, & Shahrivar Z. (2012). Delayed face recognition in children and adolescents with autism spectrum disorders. Iranian Journal of Psychiatry, 7(2), 52–56. [PubMed: 22952545]

*Tessier S, Lambert A, Scherzer P, Jemel B, & Godbout R. (2015). REM sleep and emotional face memory in typically-developing children and children with autism. Biological Psychology, 110, 107–114. doi: 10.1016/j.biopsycho.2015.07.012 [PubMed: 26219603]

*Trepagnier C, Sebrechts MM, & Peterson R. (2002). Atypical face gaze in autism. Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 5(3), 213–217. doi: 10.1089/109493102760147204

Trontel HG, Duffield TC, Bigler ED, Froehlich A, Prigge MB, Nielsen JA, … Lainhart JE (2013). Fusiform correlates of facial memory in autism. Behavioral Sciences, 3(3), 348–371. doi: 10.3390/bs3030348 [PubMed: 24761228]

Uljarevic M, & Hamilton A. (2013). Recognition of emotions in autism: a formal meta-analysis. Journal of Autism and Developmental Disorders, 43(7), 1517–1526. doi: 10.1007/s10803-012-1695-5 [PubMed: 23114566]

*Uono S, Sato W, Toichi M. (2011). The specific impairment of fearful expression recognition and its atypical development in pervasive developmental disorder. Social Neuroscience, 6(5–6), 452–463. 10.1080/17470919.2011.605593 [PubMed: 21919566]

Valentine JC, Pigott TD, & Rothstein HR (2009). How many studies do you need? Journal of Educational and Behavioral Statistics, 35(2), 215–247. doi: 10.3102/1076998609346961

Velikonja T, Fett A-K, & Velthorst E. (2019). Patterns of nonsocial and social cognitive functioning in adults with autism spectrum disorder: A systematic review and meta-analysis. JAMA Psychiatry, 76(2), 135–151. doi: 10.1001/jamapsychiatry.2018.3645 [PubMed: 30601878]

Verhallen RJ, Bosten JM, Goodbourn PT, Lawrance-Owen AJ, Bargary G, & Mollon JD (2017). General and specific factors in the processing of faces. Vision Research, 141. doi: 10.1016/j.visres.2016.12.014

Viechtbauer W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–48.URL http://www.jstatsoft.org/v36/i03/.

*Wallace S, Coleman M, & Bailey A. (2008). Face and object processing in autism spectrum disorders. Autism Research : Official Journal of the International Society for Autism Research, 1(1), 43–51. doi: 10.1002/aur.7

*Walsh JA, Creighton SE, & Rutherford. (2016). Emotion perception or social cognitive complexity: what drives face processing deficits in autism spectrum disorder? Journal of Autism and Developmental Disorders, 46(2), 615–623. doi: 10.1007/s10803-015-2606-3 [PubMed: 26439480]

Wang Y, Zhang Y, Liu L, Cui J, Wang J, Shum DH, … Chan RC (2017). A meta-analysis of working memory impairments in autism spectrum disorders. Neuropsychology Review, 27(1), 46–61. doi: 10.1007/s11065-016-9336-y [PubMed: 28102493]

Webb SJ, Neuhaus E, & Faja S. (2017). Face perception and learning in autism spectrum disorders. Quarterly Journal of Experimental Psychology (2006), 70(5), 970—986. doi: 10.1080/17470218.2016.1151059 [PubMed: 26886246]

*Webb SJ, Jones EJ, Merkle K, Murias M, Greenson J, Richards T, … Dawson G. (2010). Response to familiar faces, newly familiar faces, and novel faces as assessed by ERPs is intact in adults

with autism spectrum disorders. International Journal of Psychophysiology, 77(2), 106–117. doi: 10.1016/j.ijpsycho.2010.04.011 [PubMed: 20452382]

*Webb SJ, Merkle K, Murias M, Richards T, Aylward E, & Dawson G. (2012). ERP responses differentiate inverted but not upright face processing in adults with ASD. Social Cognitive and Affective Neuroscience, 7(5), 578–587. doi: 10.1093/scan/nsp002 [PubMed: 19454620]

Wechsler D. (1997). Wechsler Memory Scale-Third Edition (WMS-III). San Antonio, TX: The Psychological Corporation.

*Weigelt S, Koldewyn K, & Kanwisher N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. Neuroscience & Biobehavioral Reviews, 36(3), 1060–1084. doi: 10.1016/j.neubiorev.2011.12.008 [PubMed: 22212588]

*Weigelt S, Koldewyn K, & Kanwisher N. (2013). Face recognition deficits in autism spectrum disorders are both domain specific and process specific. PloS One, 8(9), e74541. doi: 10.1371/journal.pone.0074541

*White S, Hill E, Winston J, & Frith U. (2006). An islet of social ability in Asperger syndrome: judging social attributes from faces. Brain and Cognition, 61(1), 69–77. doi: 10.1016/j.bandc.2005.12.007 [PubMed: 16458402]

*Whyte EM, Behrmann M, Minshew NJ, Garcia NV, & Scherf KS (2016). Animal, but not human, faces engage the distributed face network in adolescents with autism. Developmental Science, 19(2), 306–317. doi: 10.1111/desc.12305 [PubMed: 25873084]

Whyte EM, & Scherf SK (2017). Gaze following is related to the broader autism phenotype in a sex-specific way: Building the case for distinct male and female autism phenotypes. Clinical Psychological Science, 6(2), 280–287. doi: 10.1177/2167702617738380 [PubMed: 29576931]

*Williams DL, Goldstein G, & Minshew NJ (2005). Impaired memory for faces and social scenes in autism: clinical implications of memory dysfunction. Archives of Clinical Neuropsychology, 20(1), 1–15. doi: 10.1016/j.acn.2002.08.001 [PubMed: 15620811]

Wilmer JB, Germine LT, & Nakayama K. (2014). Face recognition: a model specific ability. Frontiers in Human Neuroscience, 8, 769. doi: 10.3389/fnhum.2014.00769 [PubMed: 25346673]

*Wilson EC, Brock J, & Palermo R. (2010a). Attention to social stimuli and facial identity recognition skills in autism spectrum disorder. Journal of Intellectual Disability Research, 54(12), 1104–1115. doi: 10.1111/j.1365-2788.2010.01340.x [PubMed: 20977517]

*Wilson C, Palermo R, Brock J, Burton A. (2010b). Learning new faces in typically developing children and children on the autistic spectrum. Perception, 39(12), 1645–1658. 10.1068/p6727 [PubMed: 21425702]

*Wilson EC, Palermo R, Burton MA, & Brock J. (2011). Recognition of own- and other-race faces in autism spectrum disorders. Quarterly Journal of Experimental Psychology (2006), 64(10), 1939–1954. doi: 10.1080/17470218.2011.603052 [PubMed: 21895562]

*Wilson R, Pascalis O, & Blades M. (2007). Familiar face recognition in children with autism; the differential use of inner and outer face parts. Journal of Autism and Developmental Disorders, 37(2), 314–320. doi: 10.1007/s10803-006-0169-z [PubMed: 17378032]

*Wolf JM, Tanaka JW, Klaiman C, Cockburn J, Herlihy L, Brown C, … Schultz RT (2008). Specific impairment of face-processing abilities in children with autism spectrum disorder using the Let's Face It! skills battery. Autism Research, 1(6), 329–340. doi: 10.1002/aur.56 [PubMed: 19360688]

World Health Organization. (2004). ICD-10 : International statistical classification of diseases and related health problems: tenth revision, 2nd ed. World Health Organization.

*Yerys BE, Herrington JD, Bartley GK, Liu H-S, Detre JA, & Schultz RT (2018). Arterial spin labeling provides a reliable neurobiological marker of autism spectrum disorder. Journal of Neurodevelopmental Disorders, 10(1), 32. doi: 10.1186/s11689-018-9250-0 [PubMed: 30541425]

*Yeung MK, Lee TL, & Chan AS (2019). Impaired recognition of negative facial expressions is partly related to facial perception deficits in adolescents with high-functioning autism spectrum disorder. Journal of Autism and Developmental Disorders, 1–11. doi: 10.1007/s10803-019-03915-3 [PubMed: 30014249]

*Yi L, Fan Y, Quinn PC, Feng C, Huang D, Li J, … Lee K. (2013). Abnormality in face scanning by children with autism spectrum disorder is limited to the eye region: evidence from multi-method analyses of eye tracking data. Journal of Vision, 13(10), 5–5. doi: 10.1167/13.10.5

*Yi L, Feng C, Quinn PC, Ding H, Li J, Liu Y, & Lee K. (2014). Do individuals with and without autism spectrum disorder scan faces differently? A new multi-method look at an existing controversy. Autism Research, 7(1), 72–83. doi: 10.1002/aur.1340 [PubMed: 24124133]

*Yi L, Quinn PC, Fan Y, Huang D, Feng C, Joseph L, … Lee K. (2016). Children with autism spectrum disorder scan own-race faces differently from other-race faces. Journal of Experimental Child Psychology, 141, 177–186. doi: 10.1016/j.jecp.2015.09.011 [PubMed: 26435179]

*Yi L, Quinn PC, Feng C, Li J, Ding H, & Lee K. (2015). Do individuals with autism spectrum disorder process own- and other-race faces differently? Vision Research, 107, 124–132. doi: 10.1016/j.visres.2014.11.021 [PubMed: 25542277]

*Zaki SR, & Johnson SA (2013). The role of gaze direction in face memory in autism spectrum disorder. Autism Research, 6(4), 280–287. doi: 10.1002/aur.1292 [PubMed: 23512772]

Zhu Q, Song Y, Hu S, Li X, Tian M, Zhen Z, … Liu J. (2010). Heritability of the specific cognitive ability of face perception. Current Biology, 20(2), 137–142. doi: 10.1016/j.cub.2009.11.067 [PubMed: 20060296]

Public Significance Statement

We present a data-driven analysis of the face recognition research in autism from the last 40 years. Our findings indicate that people with ASD struggle to process information about the identity of a face. This deficit is large, such that on average 80.5% of ASD individuals perform worse than typical individuals on tests of face identity processing. This impairment likely contributes to ASD-specific difficulties with social interaction, which require the ability to identify social partners as unique individuals

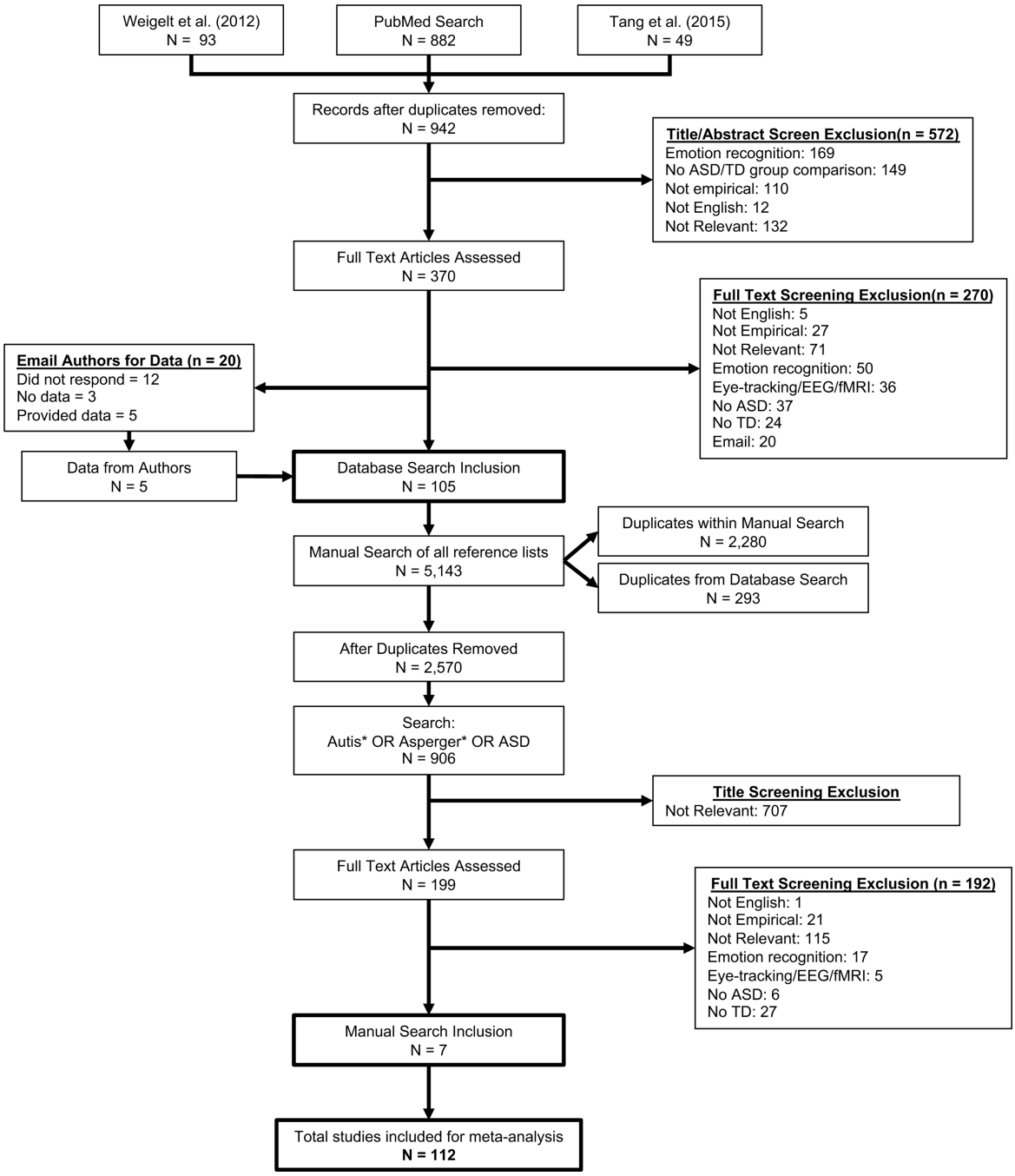

Figure 1.

PRISMA flow diagram illustrating the study identification, screening, and selection processes. ASD = autism spectrum disorder; TD = typically developing

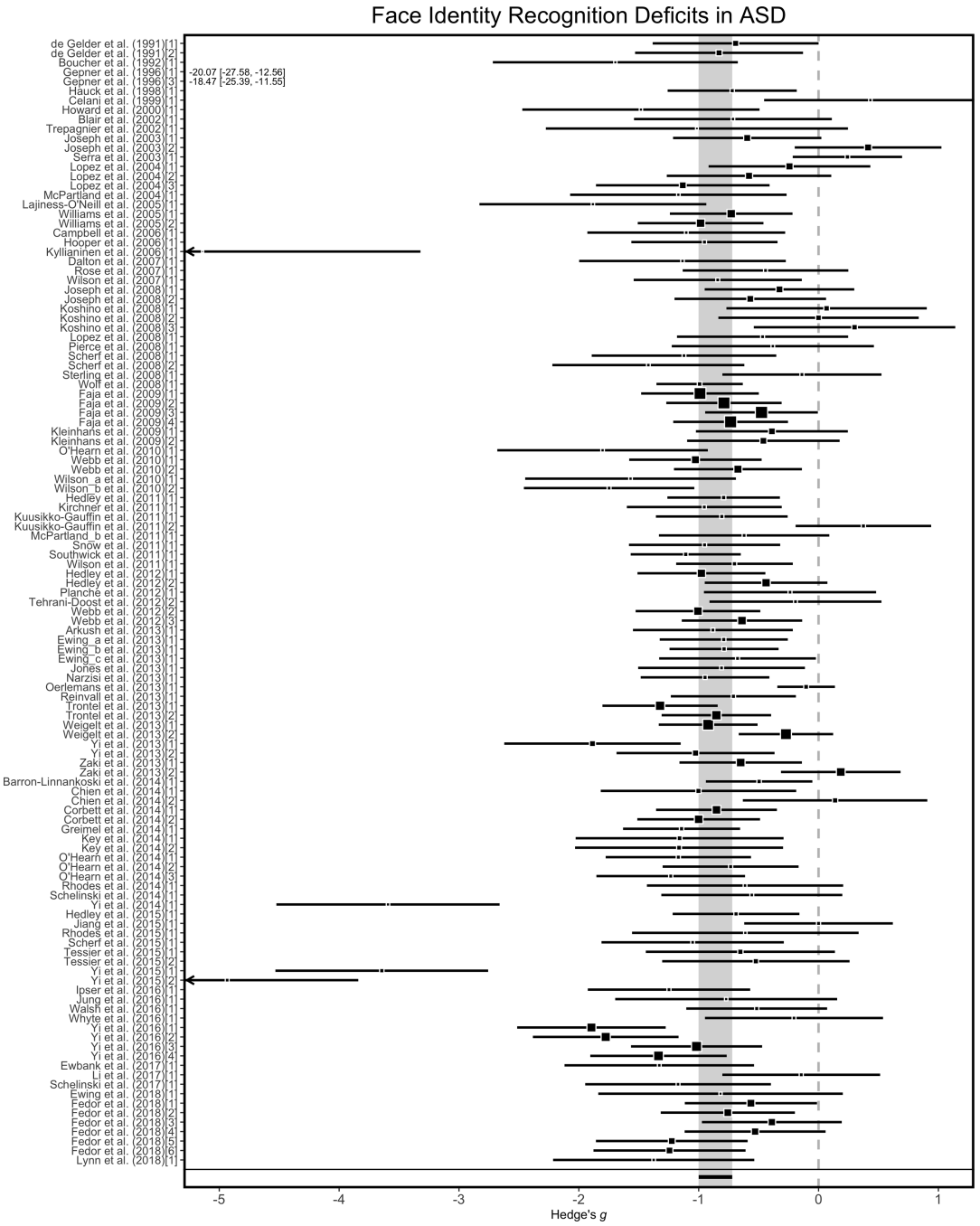

Figure 2.

Illustration of individual study and summary effect sizes for face identity recognition tasks. Forest plots containing the standardized effect size estimate (square) and 95% confidence intervals (black bar) for each effect size in the meta-analysis plotted as a function of study year and alphabetically within study year. Nested effect sizes within a study numbered in brackets. The summary effect size is presented at the bottom in black and the grey bar extending across all studies reflects the full 95% confidence interval around the summary effect size. Black arrows refer to study specific intervals that continue behind the plot area.

For effect sizes less than −5, numerical values reflect the respective Hedge's *g* and 95% CI values.

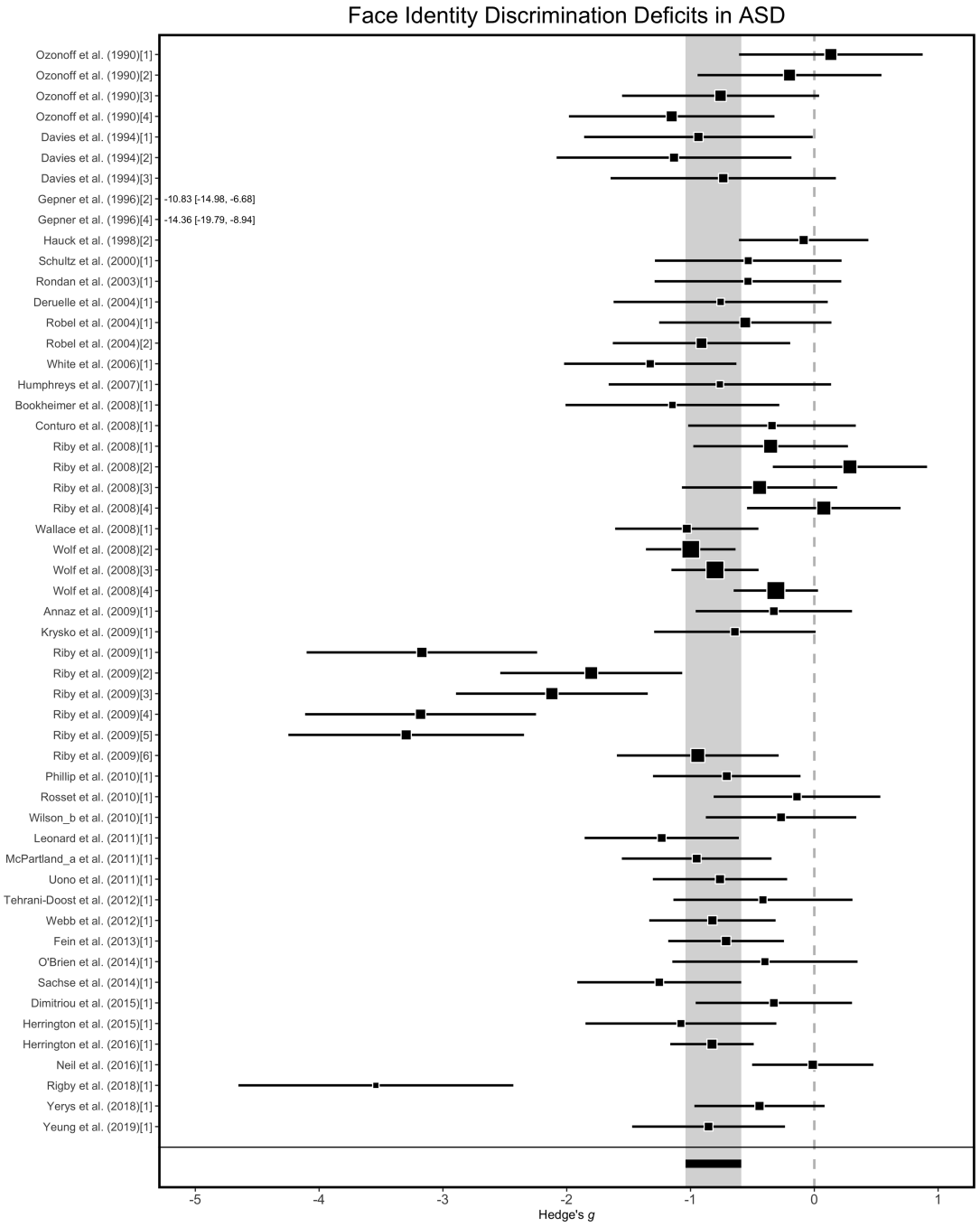

Figure 3.

Illustration of individual study and summary effect sizes for face identity discrimination tasks. Forest plots containing the standardized effect size estimate (square) and 95% confidence intervals (black bar) for each effect size in the meta-analysis plotted as a function of study year and alphabetically within study year. Nested effect sizes within a study are numbered in brackets. The summary effect size is presented at the bottom in black and the grey bar extending across all studies reflects the full 95% confidence interval around the

summary effect size. For effect sizes less than −5, numerical values reflect the respective Hedge's *g* and 95% CI values.

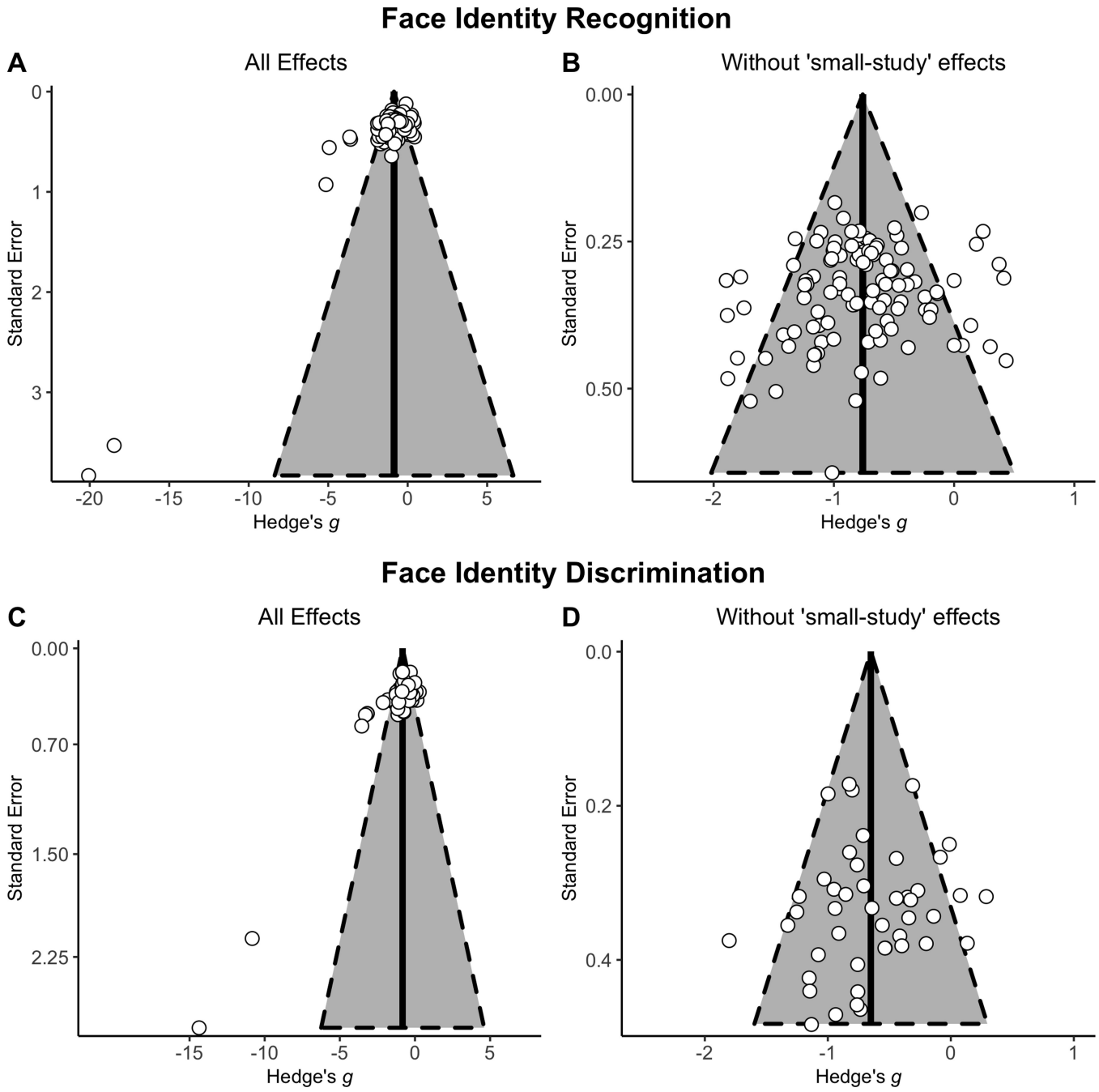

Figure 4.

Graphic illustration of small-study effects on meta-analysis. Funnel plots displaying the standardized effect size estimates (circles) as a function of precision (e.g., standard error). The funnel plot is centered on the standardized summary effect size for face identity recognition (Hedge's $g = -0.86$) and discrimination tasks (Hedge's $g = -0.82$) indicated by a vertical black line. **(A)** All face identity recognition effect sizes ($k = 119$) and **(C)** face identity discrimination effect sizes ($k = 53$) are plotted. As shown, there was significant funnel plot asymmetry for recognition and discrimination tasks with 7 and

6 effect sizes contributing to small-study effects, respectively. Upon removal, both face identity recognition **(B)** and face identity discrimination **(D)** effect sizes show symmetrical funnel plots (which indicates no presence of small-study effects). These results indicate that even after removing these studies, the summary effect size is only mildly attenuated for face recognition (Hedge's $g = -0.76$) and face discrimination (Hedge's $g = -0.65$) effect tasks.

Table 1.

Study demographics, characteristics, and effect sizes listed chronologically for face identity recognition tasks ($k = 119$).

Study	Task Name	ES description	ASD				TD				*g*	*se*
			N	Age	Sex	IQ	*N*	Age	Sex	IQ		
de Gelder et al. (1991)[1]	Old-New	Kaufman test	17	10.9 (2.3)	0.9	NA (NA)	17	8.5 (1.3)	0.9	NA (NA)	−0.7	0.4
de Gelder et al. (1991)[2]	Old-New	FACE task	17	10.9 (2.3)	0.9	NA (NA)	17	8.5 (1.3)	0.9	NA (NA)	−0.8	0.4
Boucher et al. (1992)[1]	Old-New		10	13.2 (1.8)	1.0	NA (NA)	10	13.2 (1.7)	1.0	NA (NA)	−1.7	0.5
Gepner et al. (1996)[1]	Old-New	VM-Control	7	11.2 (4.7)	0.6	NA (NA)	7	5.6 (2.4)	0.6	NA (NA)	−20.1	3.8
Gepner et al. (1996)[3]	Old-New	NVM-Control	7	11.2 (4.7)	0.6	NA (NA)	7	5.9 (1.8)	0.6	NA (NA)	−18.5	3.5
Hauck et al. (1998)[1]	Old-New		24	9.6 (1.7)	1.0	NA (NA)	34	4.7 (0.8)	0.5	NA (NA)	−0.7	0.3
Celani et al. (1999)[1]	Old-New		10	12.6 (3.7)	0.8	63.2 (19.4)	10	6.2 (1.5)	0.8	101.6 (6.6)	0.4	0.5
Howard et al. (2000)[1]	Old-New		10	NA (NA)	1.0	NA (NA)	10	NA (NA)	NA	NA (NA)	−1.5	0.5
Blair et al. (2002)[1]	Warrington		12	29.9 (7.6)	1.0	89.6 (12.2)	12	31.1 (6.8)	0.8	80.8 (13.9)	−0.7	0.4
Trepagnier et al. (2002)[1] *	Old-New		5	18.4 (NA)	0.8	NA (NA)	6	19.5 (NA)	0.7	NA (NA)	−1.0	0.6
Joseph et al. (2003)[1]	Old-New	PW-Eyes	22	10.9 (2.1)	1.0	91.0 (22.0)	20	10.8 (1.9)	0.7	91.0 (14.0)	−0.6	0.3
Joseph et al. (2003)[2]	Old-New	PW-Mouth	22	10.9 (2.1)	1.0	91.0 (22.0)	20	10.8 (1.9)	0.7	91.0 (14.0)	0.4	0.3
Serra et al. (2003)[1]	Old-New		26	8.8 (1.1)	0.8	NA (NA)	65	8.7 (1.1)	0.7	NA (NA)	0.2	0.2
Lopez et al. (2004)[1]	Old-New	PW-Eyes	17	13 (1.1)	NA	87.2 (23.3)	17	13.1 (0.1)	NA	96.5 (15.3)	−0.2	0.3
Lopez et al. (2004)[2]	Old-New	PW-Nose	17	13 (1.1)	NA	87.2 (23.3)	17	13.1 (0.1)	NA	96.5 (15.3)	−0.6	0.4
Lopez et al. (2004)[3]	Old-New	PW-Mouth	17	13 (1.1)	NA	87.2 (23.3)	17	13.1 (0.1)	NA	96.5 (15.3)	−1.1	0.4
McPartland et al. (2004)[1]	WMS	WMS-Faces I	9	21.2 (8.3)	0.9	NA (NA)	14	24.6 (6.3)	0.9	NA (NA)	−1.2	0.5
Lajiness-O'Neill et al. (2005)[1]	TOMAL		11	8.6 (2.5)	0.9	90.0 (7.1)	14	12.9 (3.5)	0.4	89.2 (5.2)	−1.9	0.5
Williams et al. (2005)[1]	WMS	Faces I	29	28.7 (10.4)	0.9	105.9 (14.2)	34	26.5 (10.2)	0.9	109.7 (11.4)	−0.7	0.3
Williams et al. (2005)[2]	WMS	Face II	29	28.7 (10.4)	0.9	105.9 (14.2)	34	26.5 (10.2)	0.9	109.7 (11.4)	−1.0	0.3
Campbell et al. (2006)[1]	Warrington		13	13.2 (1.8)	0.8	NA (NA)	13	13.3 (2.1)	0.8	NA (NA)	−1.1	0.4
Hooper et al. (2006)[1]	NEPSY		23	9.6 (2.3)	0.8	97.4 (15.5)	23	9.7 (2.3)	0.8	104.7 (10.4)	−1.0	0.3
Kyllianinen et al. (2006)[1] *	n-Back		10	9.8 (1.4)	NA	91.0 (17.0)	10	9.1 (1.2)	NA	103.0 (6.0)	−5.1	0.9
Dalton et al. (2007)[1] *	Old-New		12	14.4 (4.8)	0.7	110.0 (15.7)	12	14.2 (3.6)	0.8	115.8 (8.3)	−1.1	0.4
Rose et al. (2007)[1]	n-Back		16	10.3 (1.6)	0.8	NA (NA)	17	10.0 (1.9)	0.9	NA (NA)	−0.4	0.4
Wilson et al. (2007)[1]	Old-New		17	8.6 (NA)	0.9	NA (NA)	17	8.3 (NA)	1.0	NA (NA)	−0.8	0.4

Study	Task Name	ES description	ASD				TD				*g*	*se*
			N	Age	Sex	IQ	*N*	Age	Sex	IQ		
Joseph et al. (2008)[1]	Old-New	PW-Mouth	20	12.5 (2.0)	0.9	NA (NA)	20	11.9 (1.8)	0.6	NA (NA)	−0.3	0.3
Joseph et al. (2008)[2]	Old-New	PW-Condition	20	12.5 (2.0)	0.9	NA (NA)	20	11.9 (1.8)	0.6	NA (NA)	−0.6	0.3
Koshino et al. (2008)[1]*	n-Back	0-back	11	24.5 (10.2)	1.0	104.5 (13.1)	11	28.7 (10.9)	0.9	108.6 (9.1)	0.1	0.4
Koshino et al. (2008)[2]*	n-Back	1-back	11	24.5 (10.2)	1.0	104.5 (13.1)	11	28.7 (10.9)	0.9	108.6 (9.1)	0.0	0.4
Koshino et al. (2008)[3]*	n-Back	2-back	11	24.5 (10.2)	1.0	104.5 (13.1)	11	28.7 (10.9)	0.9	108.6 (9.1)	0.3	0.4
Lopez et al. (2008)[1]	Old-New		15	13.8 (2.3)	NA	87.1 (24.9)	16	14.3 (0.8)	NA	98.8 (16.2)	−0.5	0.4
Pierce et al. (2008)[1]*	n-Back		11	9.9 (NA)	0.8	91.0 (NA)	11	9.8 (NA)	0.8	108.5 (NA)	−0.4	0.4
Scherf et al. (2008)[1]	Old-New	Children	15	11.0 (1.1)	1.0	102.0 (15.0)	15	12.0 (1.0)	0.8	104 (7.0)	−1.1	0.4
Scherf et al. (2008)[2]	Old-New	Adults	15	32.0 (13.0)	0.9	103.0 (16.0)	15	22.0 (5.0)	0.9	111 (10.0)	−1.4	0.4
Sterling et al. (2008)[1]*	Old-New		17	23.5 (7.2)	0.9	107.1 (13.3)	18	24.2 (6.9)	0.9	109.9 (13.2)	−0.1	0.3
Wolf et al. (2008)[1]	Old-New		66	11.9 (4.0)	0.8	106.8 (20.9)	67	11.9 (3.0)	0.6	106.8 (8.0)	−1.0	0.2
Faja et al. (2009)[1]	WMS	Faces I	39	24 (7.4)	0.9	111.0 (15.0)	33	24.6 (7.1)	0.9	110 (13.0)	−1.0	0.3
Faja et al. (2009)[2]	WMS	Faces II	39	24 (7.4)	0.9	111.0 (15.0)	33	24.6 (7.1)	0.9	110 (13.0)	−0.8	0.2
Faja et al. (2009)[3]	Old-New	PW-Eyes	39	24 (7.4)	0.9	111.0 (15.0)	33	24.6 (7.1)	0.9	110.0 (13.0)	−0.5	0.2
Faja et al. (2009)[4]	Old-New	PW-Mouth	39	24 (7.4)	0.9	111.0 (15.0)	33	24.6 (7.1)	0.9	110.0 (13.0)	−0.7	0.2
Kleinhans et al. (2009)[1]*	n-Back	Run 1	19	21.9 (5.9)	NA	107.0 (13.8)	20	24.7 (7.9)	NA	110.5 (13.9)	−0.4	0.3
Kleinhans et al. (2009)[2]*	n-Back	Run 2	19	21.9 (5.9)	NA	107.0 (13.8)	20	24.7 (7.9)	NA	110.5 (13.9)	−0.5	0.3
O'Hearn et al. (2010)[1]	CFMT		14	23.1 (4.5)	NA	107.4 (12.5)	14	22.7 (4.5)	NA	105.4 (10.6)	−1.8	0.4
Webb et al. (2010)[1]	WMS	Faces I	29	22.4 (6.1)	0.9	110.2 (14.0)	28	24.0 (7.0)	0.9	111.7 (12.5)	−1.0	0.3
Webb et al. (2010)[2]	WMS	Faces II	29	22.4 (6.1)	0.9	110.2 (14.0)	28	24.0 (7.0)	0.9	111.7 (12.5)	−0.7	0.3
Wilson et al. (2010)[1]	Old-New		13	10.1 (1.9)	0.7	NA (NA)	13	10.7 (2.1)	0.5	NA (NA)	−1.6	0.4
Wilson et al. (2010B)[2]	Old-New		21	10.2 (2.3)	0.8	NA (NA)	21	7.3 (1.7)	NA	NA (NA)	−1.7	0.4
Hedley et al. (2011)[1]	CFMT		34	29.9 (11.6)	0.7	105.6 (14.9)	42	24.6 (7.4)	0.3	109.6 (9.1)	−0.8	0.2
Kirchner et al. (2011)[1]	CFMT		20	31.9 (7.6)	0.8	112.6 (11.6)	21	31.8 (7.4)	0.7	110.1 (8.7)	−1.0	0.3
Kuusikko-Gauffin et al. (2011)[1]	NEPSY	Younger Group	28	10.3 (1.1)	0.7	107.1 (15.7)	27	9.9 (1.1)	0.5	NA (NA)	−0.8	0.3
Kuusikko-Gauffin et al. (2011)[2]	NEPSY	Older Group	17	13.6 (1.6)	0.9	107.2 (9.4)	43	14 (1.2)	0.4	NA (NA)	0.4	0.3
McPartland et al. (2011b)[1]	CMS		15	14.5 (1.7)	0.9	NA (NA)	17	14.5 (1.3)	0.8	NA (NA)	−0.6	0.4
Snow et al. (2011)[1]*	Old-New		22	16.0 (2.4)	1.0	111.5 (17.6)	21	16.8 (1.9)	0.8	110.3 (10.1)	−1.0	0.3

Study	Task Name	ES description	ASD				TD				*g*	*se*
			N	Age	Sex	IQ	*N*	Age	Sex	IQ		
Southwick et al. (2011)[1]	TOMAL		50	11.6 (4.3)	1.0	NA (NA)	36	12.3 (4.2)	1.0	NA (NA)	−1.1	0.2
Wilson et al. (2011)[1]	Old-New		27	10.1 (2.2)	0.8	NA (NA)	47	8.5 (2.9)	0.5	NA (NA)	−0.7	0.2
Hedley et al. (2012)[1] *	Old-New	Same viewpoint	24	NA (NA)	NA	NA (NA)	40	NA (NA)	NA	NA (NA)	−1.0	0.3
Hedley et al. (2012)[2] *	Old-New	Viewpoint change	24	NA (NA)	NA	NA (NA)	40	NA (NA)	NA	NA (NA)	−0.4	0.3
Planche et al. (2012)[1]	NEPSY		15	8.5 (NA)	0.9	98.1 (26.2)	15	9.0 (NA)	0.8	106 (8.3)	−0.2	0.4
Tehrani-Doost et al. (2012)[2]	Old-New	Delayed Benton	15	12.8 (3.2)	NA	99.0 (11.9)	15	10.5 (3.0)	NA	113.5 (8.3)	−0.2	0.4
Webb et al. (2012)[2]	WMS	Faces I	32	23.1 (6.9)	0.9	111.3 (13.9)	32	23.7 (6.7)	0.9	110 (12.8)	−1.0	0.3
Webb et al. (2012)[3]	WMS	Faces II	32	23.1 (6.9)	0.9	111.3 (13.9)	32	23.7 (6.7)	0.9	110 (12.8)	−0.6	0.3
Arkush et al. (2013)[1]	Old-New		18	18.3 (1.4)	0.6	98.0 (12.0)	20	18.3 (1.7)	0.6	118.0 (8.0)	−0.9	0.3
Ewing et al. (2013a)[1]	CFMT		29	11.8 (2.2)	0.8	NA (NA)	29	11.8 (2.7)	0.8	NA (NA)	−0.8	0.3
Ewing et al. (2013b)[1]	Old-New		40	11.5 (2.2)	0.8	NA (NA)	40	11.6 (2.7)	0.8	NA (NA)	−0.8	0.2
Ewing et al. (2013c)[1]	CFMT		19	11.4 (2.2)	0.9	NA (NA)	19	11.4 (2.7)	0.9	NA (NA)	−0.7	0.3
Jones et al. (2013)[1]	Old-New		20	9.1 (0.9)	1.0	NA (NA)	15	9.9 (0.8)	0.6	NA (NA)	−0.8	0.4
Narzisi et al. (2013)[1]	NEPSY		22	9.8 (3.6)	1.0	99.1 (14.2)	44	NA (NA)	NA	NA (NA)	−0.9	0.3
Oerlemans et al. (2013)[1]	Old-New		140	12.4 (3.0)	0.8	102 (13.8)	127	11.0 (3.6)	0.4	107.4 (12.4)	−0.1	0.1
Reinvall et al. (2013)[1]	NEPSY		30	13.5 (1.2)	0.7	103.2 (10.7)	30	13.7 (1.0)	0.7	NA (NA)	−0.7	0.3
Trontel et al. (2013)[1]	TOMAL	Immediate	56	12.0 (4.4)	1.0	98.3 (16.6)	31	12.0 (4.0)	1.0	115.2 (15.6)	−1.3	0.2
Trontel et al. (2013)[2]	TOMAL	Delayed	56	12.0 (4.4)	1.0	98.3 (16.6)	31	12.0 (4.0)	1.0	115.2 (15.6)	−0.9	0.2
Weigelt et al. (2013)[1]	Old-New		50	9.2 (1.8)	0.9	NA (NA)	50	9.1 (1.9)	0.9	NA (NA)	−0.9	0.2
Weigelt et al. (2013)[2]	Matching		50	9.2 (1.8)	0.9	NA (NA)	50	9.1 (1.9)	0.9	NA (NA)	−0.3	0.2
Yi et al. (2013)[1] *	Old-New	AM-Control	20	7.8 (1.6)	0.8	77.2 (19.6)	21	7.7 (1.5)	0.9	89.4 (11.2)	−1.9	0.4
Yi et al. (2013)[2] *	Old-New	IM-Control	20	7.8 (1.6)	0.8	77.2 (19.6)	20	5.7 (0.8)	0.9	98.1 (7.0)	−1.0	0.3
Zaki et al. (2013)[1]	Old-New	Direct Gaze	31	13.1 (2.8)	0.9	111.7 (13.7)	31	13.1 (2.4)	0.9	111.8 (12.2)	−0.6	0.3
Zaki et al. (2013)[2]	Old-New	Averted Gaze	31	13.1 (2.8)	0.9	111.7 (13.7)	31	13.1 (2.4)	0.9	111.8 (12.2)	0.2	0.3
Barron-Linnankoski et al. (2014)[1]	NEPSY		30	9.1 (1.3)	0.9	107.2 (17.3)	60	9.1 (1.4)	0.9	NA (NA)	−0.5	0.2
Chien et al. (2014)[1]	Old-New	Asian Faces	13	7.6 (1.4)	0.8	NA (NA)	13	7.6 (1.4)	0.7	NA (NA)	−1.0	0.4
Chien et al. (2014)[2]	Old-New	African Faces	13	7.6 (1.4)	0.8	NA (NA)	13	7.6 (1.4)	0.7	NA (NA)	0.1	0.4
Corbett et al. (2014)[1]	NEPSY	Immediate	34	10.0 (NA)	1.0	100.6 (18.6)	32	9.6 (NA)	1.0	118.1 (13.8)	−0.9	0.3
Corbett et al. (2014)[2]	NEPSY	Delayed	34	10.0 (NA)	1.0	100.6 (18.6)	32	9.6 (NA)	1.0	118.1 (13.8)	−1.0	0.3

Study	Task Name	ES description	ASD				TD				*g*	*se*
			N	Age	Sex	IQ	*N*	Age	Sex	IQ		
Greimel et al. (2014)[1]	ANT		38	21.1 (9.5)	1.0	107.7 (13.2)	37	20.6 (7.0)	1.0	113 (10.2)	−1.1	0.2
Key et al. (2014)[1]	NEPSY	Immediate	13	10.8 (1.5)	0.8	103.2 (13.7)	11	10.4 (1.7)	0.8	123.5 (9.8)	−1.2	0.4
Key et al. (2014)[2]	NEPSY	Delayed	13	10.8 (1.5)	0.8	103.2 (13.7)	11	10.4 (1.7)	0.8	123.5 (9.8)	−1.2	0.4
O'Hearn et al. (2014)[1]	Old-New	Children	24	11.4 (1.6)	0.9	111.5 (10.8)	25	11.3 (1.4)	0.8	109.2 (11.7)	−1.2	0.3
O'Hearn et al. (2014)[2]	Old-New	Adolescents	25	15.3 (1.5)	0.8	106.8 (13.3)	26	15.1 (1.4)	0.8	107.7 (9.5)	−0.7	0.3
O'Hearn et al. (2014)[3]	Old-New	Adults	19	24.4 (4.7)	1.0	110.6 (14.8)	31	24.0 (5.1)	0.8	111.5 (10.4)	−1.2	0.3
Rhodes et al. (2014)[1]	CFMT		12	12.2 (1.7)	0.9	NA (NA)	12	12.4 (1.8)	0.9	NA (NA)	−0.6	0.4
Schelinski et al. (2014)[1]	CFMT		14	28.9 (7.7)	0.7	110.1 (12.2)	14	29.4 (7.1)	0.7	105.1 (13.0)	−0.6	0.4
Yi et al. (2014)[1]*	Old-New		19	20.8 (3.3)	0.7	NA (NA)	28	20.6 (2.9)	0.8	NA (NA)	−3.6	0.5
Hedley et al. (2015)[1]*	Old-New		26	28.9 (9.5)	0.6	106.6 (15.1)	33	25 .0 (10.0)	0.6	112.1 (9.3)	−0.7	0.3
Jiang et al. (2015)[1]	Old-New		20	11.4 (2.5)	0.9	NA (NA)	20	11.5 (2.3)	0.8	NA (NA)	0.0	0.3
Rhodes et al. (2015)[1]	CFMT		9	11.1 (2.0)	1.0	NA (NA)	9	11.3 (2.8)	0.4	NA (NA)	−0.6	0.5
Scherf et al. (2015)[1]	CFMT		20	14.1 (NA)	1.0	108.5 (NA)	12	13.8 (NA)	1.0	117.6 (NA)	−1.1	0.4
Tessier et al. (2015)[1]	Old-New	Immediate	13	10.2 (2.1)	1.0	105.2 (19.8)	13	10.2 (2.0)	1.0	116.5 (10.3)	−0.7	0.4
Tessier et al. (2015)[2]	Old-New	Delayed	13	10.2 (2.1)	1.0	105.2 (19.8)	13	10.2 (2.0)	1.0	116.5 (10.3)	−0.5	0.4
Yi et al. (2015)[1]*	Old-New	Own race	24	20.7 (3.9)	0.7	NA (NA)	28	20.6 (3.3)	0.8	NA (NA)	−3.6	0.5
Yi et al. (2015)[2]*	Old-New	Other race	24	20.7 (3.9)	0.7	NA (NA)	28	20.6 (3.3)	0.8	NA (NA)	−4.9	0.6
Ipser et al. (2016)[1]	Old-New		20	43.2 (12.3)	0.8	108.7 (15.2)	20	44 (13.9)	0.7	110.8 (13.9)	−1.2	0.3
Jung et al. (2016)[1]*	n-Back		8	15.6 (9.6)	1.0	95.9 (14.9)	12	14.5 (10.8)	1.0	112.2 (12.4)	−0.8	0.5
Walsh et al. (2016)[1]	Old-New		23	30.8 (8.5)	0.8	97.4 (11.0)	23	28.4 (9.3)	0.8	97.4 (13.1)	−0.5	0.3
Whyte et al. (2016)[1]	CFMT		14	15.0 (2.0)	0.9	112.0 (11.0)	14	15.0 (2.0)	0.9	110 (12)	−0.2	0.4
Yi et al. (2016)[1]*	Old-New	AM-Control/Own race	29	7.9 (1.4)	0.9	NA (NA)	29	7.9 (1.4)	0.9	NA (NA)	−1.9	0.3
Yi et al. (2016)[2]*	Old-New	AM-Control/ Other race	29	7.9 (1.4)	0.9	NA (NA)	29	7.9 (1.4)	0.9	NA (NA)	−1.8	0.3
Yi et al. (2016)[3]*	Old-New	IM-Control/Own race	29	7.9 (1.4)	0.9	NA (NA)	29	5.7 (1.0)	0.9	NA (NA)	−1.0	0.3
Yi et al. (2016)[4]*	Old-New	IM-Control/Other race	29	7.9 (1.4)	0.9	NA (NA)	29	5.7 (1.0)	0.9	NA (NA)	−1.3	0.3
Ewbank et al. (2017)[1]	CFMT		15	31.8 (9.2)	0.7	126.0 (11.8)	15	28.1 (7.5)	0.6	128.2 (10.5)	−1.3	0.4

Study	Task Name	ES description	ASD				TD				*g*	*se*
			N	Age	Sex	IQ	*N*	Age	Sex	IQ		
Li et al. (2017)[1]	Old-New		16	6.0 (0.7)	0.9	NA (NA)	20	5.8 (0.5)	1.0	NA (NA)	−0.1	0.3
Schelinski et al. (2017)[1]	CFMT		15	33.7 (10.4)	0.9	110.5 (14.2)	15	33.8 (9.7)	0.9	110 (11.4)	−1.2	0.4
Ewing et al. (2018)[1]	CFMT		8	10.6 (1.2)	0.6	NA (NA)	8	8.9 (0.8)	0.4	NA (NA)	−0.8	0.5
Fedor et al. (2018)[1]*	CFMT	Children	24	11.2 (1.7)	0.9	112.8 (12.9)	29	11.4 (1.4)	0.8	107.5 (12.6)	−0.6	0.3
Fedor et al. (2018)[2]*	CFMT	Children	24	11.2 (1.7)	0.9	112.8 (12.9)	29	11.4 (1.4)	0.8	107.5 (12.6)	−0.8	0.3
Fedor et al. (2018)[3]*	CFMT	Adolescents	23	15.3 (1.5)	0.8	107.7 (13.3)	23	15.4 (1.5)	0.8	108.6 (8.5)	−0.4	0.3
Fedor et al. (2018)[4]*	CFMT	Adolescents	23	15.3 (1.5)	0.8	107.7 (13.3)	23	15.4 (1.5)	0.8	108.6 (8.5)	−0.5	0.3
Fedor et al. (2018)[5]*	CFMT	Adults	19	24.2 (4.9)	1.0	111.3 (14.6)	28	24 (5.3)	0.8	112.1 (11.7)	−1.2	0.3
Fedor et al. (2018)[6]*	CFMT	Adults	19	24.2 (4.9)	1.0	111.3 (14.6)	28	24 (5.3)	0.8	112.1 (11.7)	−1.2	0.3
Lynn et al. (2018)[1]*	CFMT	Adults	13	24.9 (4.9)	NA	115.3 (13.6)	14	24.1 (5.5)	NA	114.9 (11.0)	−1.4	0.4

Note. Study authors are denoted by the first author, year, and unique effect size identifier in brackets. For studies with multiple effect sizes, the ES description column differentiates multiple effect sizes. For descriptive statistics, NA indicate data was not available in the article. VM-Control = Verbal-matched control group; NVM-Control = Nonverbal -matched control group; AM-Control = Age-matched control group; IM-Control = IQ-matched control group; PW-Eyes = Parts/Whole task identity change of eyes; PW-Mouth = Parts/Whole task identity change of mouth; ASD = Autism Spectrum Disorder, TD = typically developing, CFMT = Cambridge Face Memory Test, NEPSY = A Developmental Neuropsychological Assessment, TOMAL = Test of Memory and Learning, WMS = Weschler Memory Scale, Sex = % male; *g* = standardized effect size (Hedge's *g*); *se* = standard error.

* indicates that the outcome measure was collected either in the MRI scanner or during eye-tracking.

Table 2.

Study demographics, characteristics, and effect sizes listed chronologically for face identity discrimination tasks ($k = 53$).

Study	Task	Effect Size Description	ASD				TD				*g*	*se*
			N	Age	Sex	IQ	*N*	Age	Sex	IQ		
Ozonoff et al. (1990)[1]	Sorting	Exp. 1	14	6.4 (2.0)	0.7	NA (NA)	14	3.0 (0.3)	0.7	NA (NA)	0.1	0.4
Ozonoff et al. (1990)[2]	Matching	Exp. 1	14	6.4 (2.0)	0.7	NA (NA)	14	3.0 (0.3)	0.7	NA (NA)	−0.2	0.4
Ozonoff et al. (1990)[3]	Sorting	Exp. 2	13	6.2 (2.1)	0.7	76.0 (20.0)	13	4.1 (1.5)	0.7	114.0 (20.0)	−0.8	0.4
Ozonoff et al. (1990)[4]	Matching	Exp. 2	13	6.2 (2.1)	0.7	76.0 (20.0)	13	4.1 (1.5)	0.7	114.0 (20.0)	−1.2	0.4
Davies et al. (1994)[1]	Matching		10	14.9 (1.9)	NA	NA (NA)	10	14.7 (2.1)	NA	NA (NA)	−0.9	0.5
Davies et al. (1994)[2]	Sorting	Viewpoint change	9	14.3 (1.5)	NA	NA (NA)	11	13.9 (1.4)	NA	NA (NA)	−1.1	0.5
Davies et al. (1994)[3]	Sorting	Expression change	9	14.3 (1.5)	NA	NA (NA)	11	13.9 (1.4)	NA	NA (NA)	−0.7	0.5
Gepner et al. (1996)[2]	Sorting	VM-Control	7	11.2 (4.7)	0.6	NA (NA)	7	5.6 (2.4)	0.6	NA (NA)	−10.8	2.1
Gepner et al. (1996)[4]	Sorting	NVM-Control	7	11.2 (4.7)	0.6	NA (NA)	7	5.9 (1.8)	0.6	NA (NA)	−14.4	2.8
Hauck et al. (1998)[2]	Matching		24	9.6 (1.7)	1.0	NA (NA)	34	4.7 (0.8)	0.5	NA (NA)	−0.1	0.3
Schultz et al. (2000)[1] *	Same-Different		14	23.8 (12.4)	1.0	109.1 (19.5)	14	21.7 (7.2)	1.0	110.4 (17.2)	−0.5	0.4
Rondan et al. (2003)[1]	Matching		14	10.1 (2.5)	0.6	NA (NA)	14	10.1 (2.3)	0.6	NA (NA)	−0.5	0.4
Deruelle et al. (2004)[1]	Matching		11	9.2 (2.2)	0.6	NA (NA)	11	9.4 (2.3)	0.6	NA (NA)	−0.8	0.4
Robel et al. (2004)[1]	Same-Different	Viewpoint change	14	8.2 (1.7)	0.9	NA (NA)	20	7.9 (1.3)	0.8	NA (NA)	−0.6	0.4
Robel et al. (2004)[2]	Same-Different	Expression change	14	8.2 (1.7)	0.9	NA (NA)	20	7.9 (1.3)	0.8	NA (NA)	−0.9	0.4
White et al. (2006)[1]	Benton		16	32.3 (14.2)	0.6	111.8 (15.6)	24	37.8 (12.4)	0.5	115.2 (11.3)	−1.3	0.4
Humphreys et al. (2007)[1]	Benton		18	24.7 (9.0)	0.9	103.9 (16.1)	7	24.4 (6.5)	1.0	110.6 (9.0)	−0.8	0.5
Bookheimer et al. (2008)[1] *	Matching		12	11.3 (4.0)	1.0	NA (NA)	12	11.9 (2.4)	1.0	NA (NA)	−1.1	0.4
Conturo et al. (2008)[1]	Benton		17	26.5 (11.3)	0.8	104.4 (8.6)	17	26.1 (11.1)	0.8	105.2 (9.7)	−0.3	0.3
Riby et al. (2008)[1]	Matching	VM-Control	20	12.0 (2.8)	0.8	NA (NA)	20	7.5 (1.0)	0.6	NA (NA)	−0.4	0.3
Riby et al. (2008)[2]	Matching	VMA-Control/Inner face	20	12.0 (2.8)	0.8	NA (NA)	20	7.5 (1.0)	0.6	NA (NA)	0.3	0.3
Riby et al. (2008)[3]	Matching	NVM-Control	20	12.0 (2.8)	0.8	NA (NA)	20	8.8 (2.5)	0.7	NA (NA)	−0.4	0.3
Riby et al. (2008)[4]	Matching	NVM-Control/Inner face	20	12.0 (2.8)	0.8	NA (NA)	20	8.8 (2.5)	0.7	NA (NA)	0.1	0.3
Wallace et al. (2008)[1]	Benton		26	32.0 (9.0)	0.9	101.0 (18.0)	26	31.0 (9.0)	0.9	98.0 (12.0)	−1.0	0.3
Wolf et al. (2008)[2]	Matching		66	11.7 (3.8)	0.8	107.0 (20.5)	66	11.7 (3.0)	0.6	107.0 (8.0)	−1.0	0.2
Wolf et al. (2008)[3]	Matching	PW-Eyes	66	11.9 (4.0)	0.8	106.8 (20.9)	68	11.9 (3.1)	0.6	106.8 (7.8)	−0.8	0.2

Study	Task	Effect Size Description	ASD				TD				*g*	*se*
			N	Age	Sex	IQ	*N*	Age	Sex	IQ		
Wolf et al. (2008)[4]	Matching	PW-Mouth	66	11.9 (4.0)	0.8	106.8 (20.9)	68	11.9 (3.1)	0.6	106.8 (7.8)	−0.3	0.2
Anz et al. (2009)[1]	Benton		16	8.4 (1.8)	0.8	NA (NA)	25	7.2 (2.8)	0.5	NA (NA)	−0.3	0.3
Krysko et al. (2009)[1]	Same-Different		19	29.3 (9.4)	1.0	94.8 (12.7)	19	28.6 (6.9)	1.0	NA (NA)	−0.6	0.3
Riby et al. (2009)[1]	Matching	Upper face	20	14.8 (2.4)	0.8	NA (NA)	20	14.9 (2.2)	0.7	NA (NA)	−3.2	0.5
Riby et al. (2009)[2]	Matching	Lower face	20	14.8 (2.4)	0.8	NA (NA)	20	14.9 (2.2)	0.7	NA (NA)	−1.8	0.4
Riby et al. (2009)[3]	Matching	Featural change	20	14.8 (2.4)	0.8	NA (NA)	20	14.9 (2.2)	0.7	NA (NA)	−2.1	0.4
Riby et al. (2009)[4]	Matching	Configural change	20	14.8 (2.4)	0.8	NA (NA)	20	14.9 (2.2)	0.7	NA (NA)	−3.2	0.5
Riby et al. (2009)[5]	Matching	PW-Eyes	20	14.8 (2.4)	0.8	NA (NA)	20	14.9 (2.2)	0.7	NA (NA)	−3.3	0.5
Riby et al. (2009)[6]	Matching	PW-Mouth	20	14.8 (2.4)	0.8	NA (NA)	20	14.9 (2.2)	0.7	NA (NA)	−0.9	0.3
Phillip et al. (2010)[1]	Benton		23	32.5 (10.9)	0.7	101.5 (18.5)	23	32.4 (11.1)	0.7	112.2 (8.5)	−0.7	0.3
Rosset et al. (2010)[1]	Same-Different		17	10.5 (2.8)	0.8	93.0 (16.0)	17	10.5 (2.7)	0.8	NA (NA)	−0.1	0.3
Wilson et al. (2010b)[1]	Matching		21	10.2 (2.3)	0.8	NA (NA)	21	7.3 (1.7)	NA	NA (NA)	−0.3	0.3
Leord et al. (2011)[1]	Benton		17	10.3 (2.4)	0.9	NA (NA)	36	11.4 (2.6)	0.5	NA (NA)	−1.2	0.3
McPartland et al. (2011a)[1]	Benton		36	11.2 (3.4)	0.9	105.2 (17.3)	17	12.6 (2.4)	0.9	112.9 (13.4)	−1.0	0.3
Uono et al. (2011)[1]	Benton		28	17.6 (5.2)	0.8	103.3 (13.1)	28	18.0 (4.0)	0.9	NA (NA)	−0.8	0.3
Tehrani-Doost et al. (2012)[1]	Benton		15	12.8 (3.2)	NA	99.0 (11.9)	15	10.5 (3.0)	NA	113.5 (8.3)	−0.4	0.4
Webb et al. (2012)[1]	Benton		32	23.1 (6.9)	0.9	111.3 (13.9)	32	23.7 (6.7)	0.9	110 (12.8)	−0.8	0.3
Fein et al. (2013)[1]	Benton		43	NA (NA)	NA	NA (NA)	33	NA (NA)	NA	NA (NA)	−0.7	0.2
O'Brien et al. (2014)[1]	Benton		14	33.9 (NA)	0.8	NA (NA)	14	31.1 (NA)	0.5	NA (NA)	−0.4	0.4
Sachse et al. (2014)[1]	Benton		22	20.9 (5.6)	0.8	100.1 (13.3)	20	20.1 (3.8)	0.8	105.4 (10.5)	−1.3	0.3
Dimitriou et al. (2015)[1]	Benton		16	8.4 (1.8)	0.8	NA (NA)	25	7.2 (2.8)	0.5	NA (NA)	−0.3	0.3
Herrington et al. (2015)[1]	Benton		12	13.4 (4.2)	NA	108.8 (14.8)	19	13.4 (3.5)	NA	114.8 (10.8)	−1.1	0.4
Herrington et al. (2016)[1]*	Same-Different		81	12.5 (2.6)	0.8	101.0 (23.0)	67	12.5 (2.7)	0.7	115 (16.1)	−0.8	0.2
Neil et al. (2016)[1]	Sorting		32	11.1 (2.6)	0.8	98.6 (14.8)	32	10.6 (2.2)	0.6	101.5 (11.4)	0.0	0.3
Rigby et al. (2018)[1]	Same-Different		16	27.8 (7.8)	0.7	106.3 (10.8)	16	27.3 (7.5)	0.7	113.4 (8.7)	−3.5	0.6
Yerys et al. (2018)[1]	Benton		33	14.9 (1.7)	1.0	94.0 (20.0)	25	14.9 (1.7)	1.0	123.0 (18.0)	−0.4	0.3
Yeung et al. (2019)[1]	Benton		22	14.4 (2.2)	0.9	104.2 (15.6)	22	14.3 (1.8)	0.7	106.7 (12.9)	−0.9	0.3

Note. Study authors are denoted by the first author, year, and unique effect size identifier in brackets. For studies with multiple effect sizes, the ES description column differentiates multiple effect sizes. For descriptive statistics, NA indicate data was not available in the article. VM-Control = Verbal-matched control group; NVM-Control = Nonverbal -matched control group; PW-Eyes = Parts/Whole task

identity change of eyes; PW-Mouth = Parts/Whole task identity change of mouth; ASD = Autism Spectrum Disorder, TD = typically developing, Benton = Benton Face Recognition Test, Sex = % male; *g* = standardized effect size (Hedge's *g*); *se* = standard error.

* indicates that the outcome measure was collected either in the MRI scanner, or during eye-tracking.