Neuroscience and Biobehavioral Reviews 133 (2022) 104518

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

A systematic review and meta-analysis of facial emotion recognition in autism spectrum disorder: The specificity of deficits and the role of task characteristics

Michael K. Yeung [a,b,*]

[a] Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada
[b] Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

ARTICLE INFO

Keywords:
Autism spectrum disorder
Facial emotion recognition
Facial perception
Basic emotions
Meta-analysis

ABSTRACT

This review assessed the specificity of facial emotion recognition impairment and the role of task characteristics in facial emotion recognition in autism spectrum disorder (ASD). Based on subsets of 148 studies identified in PubMed and PsycINFO, random-effects meta-analyses showed significant impairment in recognizing all basic facial emotions in ASD. Additionally, ASD involves poorer facial emotion recognition than other clinical conditions and has similar impairment in recognizing emotional and nonemotional facial attributes, as well as in recognizing emotion in faces and other modalities. Furthermore, there are significant moderating effects for emotion complexity and holistic processing, a statistical trend for task type, and no significant effect for motion, social relevance, or stimulus salience on facial emotion recognition in ASD. Altogether, this review suggests nonselective facial emotion recognition impairment in ASD. Such impairment is relatively specific to ASD but is not specific to the recognition of emotional facial attributes or emotion recognition in the face modality. Identifying the role of task characteristics improves our understanding of the mechanisms underlying facial emotion recognition in ASD.

1. Introduction

Facial expressions convey important information regarding the emotional state of others (Darwin, 1872). Thus, the ability to recognize facial emotions is crucial for social competence (Izard et al., 2001). One of the defining features of autism spectrum disorder (ASD) is social communication and interaction difficulties (American Psychiatric Association, 2013). While many studies have reported impairment in facial emotion recognition in ASD, some studies have failed to report facial emotion recognition deficits or reported the impaired recognition of only certain basic emotions in this disorder (Harms et al., 2010). Despite these mixed findings, one meta-analysis of 48 studies on the ability to recognize emotions in the visual modality (primarily faces) has demonstrated that the general ability to recognize emotions is moderately impaired in ASD, even after correcting for publication bias (Hedges' $g = 0.41$; Uljarevic and Hamilton, 2013). This meta-analysis also found significant impairment in recognizing all basic emotions except happiness, and marginally significantly greater impairment in recognizing fear than happiness in ASD. Additionally, another meta-analysis of 43 studies has also reported moderate overall impairment in recognizing basic facial emotions in ASD ($r = 0.36$; Lozier et al., 2014). However, this meta-analysis observed significant impairment in recognizing happy, angry, fearful, and surprised expressions—but not sad or disgusted expressions—in ASD after Bonferroni correction. Thus, controversy remains regarding the pattern of facial emotion recognition deficits in ASD. In the present article, the term "impairment" specifically refers to poor performance on lab-based measures rather than to real-life problems.

Although these two meta-analyses have identified moderate impairment in (facial) emotion recognition in ASD, the specificity of such impairment remains largely unclear. Some systematic reviews have identified facial emotion recognition deficits in various child and adolescent psychiatric disorders other than ASD (Collin et al., 2013), such as attention-deficit/hyperactivity disorder (ADHD; Bora and Pantelis, 2016). Although social communication and interaction impairment is a diagnostic characteristic of ASD, it is uncertain whether ASD has greater facial emotion recognition impairment than other clinical populations. Furthermore, the recognition of emotion from facial

* Correspondence to: Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.
E-mail address: kin-chung-michael.yeung@polyu.edu.hk.

https://doi.org/10.1016/j.neubiorev.2021.104518
Received 20 April 2020; Received in revised form 2 June 2021; Accepted 29 December 2021
Available online 31 December 2021

expressions involves interactions among perceptual, emotional, and linguistic processes, as well as their underlying brain regions, including the fusiform face area, amygdala, and inferior frontal gyrus, respectively (Adolphs, 2002; Brosch et al., 2010; Goldman and Sripada, 2005; Barrett et al., 2007). Specifically, from the information-processing perspective, facial emotion recognition involves the perceptual processing of facial stimuli followed by the retrieval of conceptual knowledge of the emotion signed by the face (Adolphs, 2002). Also, the psychological and neurological mechanisms underlying emotion recognition from facial expressions share similarities with those underlying emotion recognition in other modalities, such as voices (Schirmer and Adolphs, 2017). Thus, poor facial emotion recognition in ASD could be attributable to impairment in the visuoperceptual processing of faces rather than in the emotional processing of faces per se. It may also reflect deficits in recognizing emotions, irrespective of modality, rather than in faces specifically.

There is substantial heterogeneity in the findings of facial emotion recognition in ASD (Harms et al., 2010; Uljarevic and Hamilton, 2013). This variation may be attributable to differences in task characteristics across studies. According to a narrative review of facial emotion recognition in ASD (Harms et al., 2010), autistic individuals (i.e., people who have been diagnosed with ASD) may have greater difficulties in recognizing complex emotions, such as embarrassment, than basic emotions (Baron-Cohen et al., 1997). These individuals may also have greater deficits when facial emotion processing is made more difficult by reducing the salience of facial stimuli, such as shortening stimulus presentation duration (Clark et al., 2008) and attenuating emotion intensity (Law Smith et al., 2010). Additionally, autistic individuals may have poorer performance on tasks that can be supported by cognitive or linguistic strategies than on tasks that cannot (Piggot et al., 2004). Since autistic individuals experience local processing bias (Happé and Frith, 2006), these individuals may also have poorer performance on tasks emphasizing holistic processing, such as those using upright, low-pass filtered, and/or full-face facial stimuli, than on tasks focusing on feature-based processing, such as those that use inverted, high-pass filtered, and/or eyes-only facial stimuli (Kätsyri et al., 2008). In contrast, according to the review by Harms et al. (2010), motion (Gepner et al., 2001) and familiarity (Gross, 2004) do not appear to moderate facial emotion recognition performance in ASD.

While various task factors have been proposed to exert or not to exert a moderating effect on facial emotion recognition in ASD (Harms et al., 2010), a systematic examination of these factors is required to verify the presence or absence of their moderating effects. Additionally, while previous meta-analyses have reported emotion recognition impairment in the visual modality (Uljarevic and Hamilton, 2013) or with faces (Lozier et al., 2014) in ASD, the specificity of such impairment concerning ASD diagnosis, the emotional attributes of faces, and emotion recognition in the face modality remains largely unclear. Thus, the present review aimed to conduct qualitative and quantitative syntheses of facial emotion recognition studies comparing ASD and typically developing (TD) controls and to assess the specificity of facial emotion recognition impairment, if any, as well as the role of task characteristics in facial emotion recognition in ASD. The specificity of deficits and the role of task demands were primarily assessed by taking advantage of within-subject design studies. Evidence was gathered from studies comparing multiple levels of at least one within-subject factor between ASD and controls. Consequently, the moderating effect of each factor could be studied with minimal confounding effects from other factors. This review can improve our understanding of the underlying mechanisms of facial emotion recognition in ASD in order to inform intervention approaches to improve the emotion recognition ability—and hence social competence—of autistic individuals. The results will shed light on the possibility of using measures of facial emotion recognition as a marker for ASD.

2. Method

This study conformed to standard methodological guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Moher et al., 2009). It was not registered in a registry of systematic reviews. Study selection, data extraction, and quality assessment were carried out by the author, and the codes were checked by a second individual to ensure accuracy and objectivity (study inclusion and data extraction: V. W. C.; quality assessment: K. K. C.).

2.1. Search strategy and study selection

To identify studies on facial emotion recognition in ASD, a literature search using PubMed and PsycINFO was performed by the author on February 24, 2020. The keywords used were "autis*" or "ASD" or "Asperger" or "PDD" or "pervasive developmental disorder" and "face" or "facial" and "emotion" or "affect" or "expression" or "happ*" or "sad*" or "ang*" or "disgust*" or "surprise*" or "fear*" (see Appendix A for the complete strings). These keywords were based on those used in previous meta-analyses (Lozier et al., 2014; Uljarevic and Hamilton, 2013), and the search did not involve librarians. No limit was set for the year of publication. After the initial search, the titles and abstracts of the articles were screened.

An article was included if it met all of the following criteria: (a) an original research article published in English within a peer-reviewed journal, (b) included a group of ASD participants without known causes (e.g., Fragile X syndrome), (c) included a TD group, (d) used at least one task that asked participants to choose between at least two facial emotions or to rate how much a face belonged to an emotion category, (e) involved a task condition that only involved whole, real human faces expressing distinct basic emotions, and (f) reported qualitative (i.e., presence of impairment) or quantitative (i.e., descriptive statistics such as means and *SD*s, or inferential statistics such as *t*- and *F*-values) performance data specifically for the facial emotion recognition task. These inclusion criteria were established to ensure that the included studies allowed qualitative or quantitative assessments of facial emotion recognition ability in ASD compared to TD controls. Additionally, an article was excluded if it met any of the following criteria: (a) did not confirm an ASD diagnosis in all individuals in ASD group, (b) studied the ensemble perception of facial expressions, or (c) used a facial emotion recognition task that could be solved solely by perceptual matching, which bypassed the need to process emotions. These exclusion criteria were established to ensure that all included studies focused on the ASD population and the recognition of isolated facial expressions. Additional studies were also identified from previous meta-analyses (Lozier et al., 2014; Uljarevic and Hamilton, 2013).

2.2. Data extraction and coding

The following information from each identified article was extracted and entered into a predesigned table: author name(s), year of publication, within- and between-subject factor(s) examined, sample characteristics (i.e., sample size and mean age), sample matching variables (i.e., age, sex, and IQ), primary task characteristics, basic emotions included, dependent measure, and primary task results. Moreover, for studies that investigated at least one within- and between-subject factor (see below), information regarding the tasks used for different levels of a factor was extracted and entered into the table dedicated to that factor.

For studies that examined at least one within- or between-subject factor, the different levels of the factor were coded as follows: (a) emotion type (i.e., happiness, sadness, anger, fear, disgust, surprise, or neutral expression), (b) specificity of diagnosis (ASD, non-ASD clinical population), (c) emotion specificity (recognition of facial emotions or the nonemotional attributes of faces, (d) face specificity (recognition of emotions in faces or in other modalities), (e) task type (presence or

absence of verbal cues), (f) motion (static or dynamic faces), (g) emotion complexity (basic or complex emotions), (h) holistic processing, including the orientation (e.g., upright or inverted faces), integrity (whole faces or eyes only), spatial frequency (unfiltered, low-pass filtered, or high-pass filtered faces), and hybridity (full emotional faces or half emotional and half neutral faces) of expressions, (i) social relevance, including the humanity (real human, animal, or cartoon faces), gaze direction (direct or averted gaze), ethnicity (same or different ethnicity), and familiarity (familiar or unfamiliar faces) of expressions, and (j) stimulus salience, including stimulus presentation duration (e.g., fast or slow presentation) and emotion intensity level (e.g., high or low intensity).

The task types were defined in five categories. A labeling task involved the assignment of a word selected out of a list of emotion words to a face. A matching task involved the selection of a face among face images to be paired to another face. A discrimination task involved choosing (at least) one face that corresponded to a nonface target (e.g., a word) upon presentation of at least two faces. A judgment task involved the selection of at least two faces that depicted the same expression among a group of face images. A rating task involved the grading of how much a face corresponded to an emotion category. While labeling, discrimination, and rating tasks involve language or verbal cues, matching and judgment tasks do not.

To respect the assumption of independent observations and reduce the heterogeneity of the results, several procedures were performed. First, when two ASD subgroups (e.g., autistic disorder and Asperger's syndrome) were present in a study, they were combined into one single group by calculating the weighted mean and standard deviation. A similar approach was also taken for studies with two non-ASD clinical groups. Second, most studies have used a labeling task. Thus, when two TD groups matched for age or IQ, or for verbal or nonverbal IQ, were present in a study, the group matched for (verbal) IQ was chosen as the control group. The two TD groups were not combined so that the ASD and control groups were matched for (verbal) mental ability as closely as possible. Third, where performance data were reported separately for the task conditions, labeling tasks and tasks involving humans, direct gaze, and static, upright, or prototypical facial expressions were chosen over other task types and tasks involving nonhumans, averted gaze, and dynamic, inverted, or mild facial expressions for analysis. If more than one labeling task was used, the one including a wider range of emotions was chosen.

2.3. Statistical analysis

Qualitative and quantitative assessments were performed on the included studies. First, whether autistic individuals are poorer than TD individuals at recognizing facial emotions in general and in each emotion was examined. Then, the specificity of facial emotion recognition impairment in ASD, if any, was assessed by analyzing studies contrasting ASD with other clinical populations, the ability to recognize emotional versus nonemotional facial attributes in ASD, and the ability to recognize emotion in faces versus other modalities in ASD. Finally, the role of task characteristics in facial emotion recognition in ASD was assessed by analyzing studies contrasting differences between ASD and TD groups on verbal versus nonverbal tasks, tasks using static versus dynamic faces, basic versus complex emotion tasks, tasks emphasizing holistic versus featural processing, tasks using socially relevant versus irrelevant faces, and tasks using salient versus non-salient faces.

For all studies, the dependent measure used was accuracy. However, if a measure considering response bias, both accuracy and reaction time (RT), or discrimination threshold was used, this measure would be used. Also, if the ceiling effect—operationalized by the sum of mean accuracy and its standard deviation greater than one across conditions in either ASD or TD group—was present for accuracy, RT was adopted.

2.3.1. Qualitative assessment

For each section, a qualitative assessment was first performed on studies that classified task performance in ASD into "impaired" or "unimpaired". The rationale for this assessment was as follows: a) to review the literature exhaustively, including studies that have not allowed effect size calculation, and b) to demonstrate how many studies not providing the information necessary for meta-analysis (i.e., effect size calculation) have reported null results or were missing from the subsequent meta-analysis. The qualitative assessment was based on the results of independent-sample *t*-tests or ANOVA comparing ASD and TD groups. If a study used ANCOVA to control for covariates such as age and IQ or used Mann–Whitney *U* tests to deal with the non-normal distribution of task variables, the qualitative assessment would instead be based on the *p*-value of these tests. The significance level was set at 0.05, and Bonferroni correction was applied. Accordingly, the facial emotion recognition ability of ASD was significantly impaired if the *p*-value of a statistical test comparing ASD and TD groups was smaller than the Bonferroni-corrected *p*-value threshold (e.g., 0.025 for two tests). Additionally, the ability of ASD was marginally significantly impaired if the *p*-value of a statistical test comparing ASD and TD groups was smaller than 0.05 but larger than the Bonferroni-corrected *p*-value threshold.

2.3.2. Quantitative assessment

Then, a random-effects meta-analysis using the Hartung-Knapp-Sidik-Jonkman method (Hartung and Knapp, 2001; Sidik and Jonkman, 2002) was performed on studies that have allowed effect size calculation for each factor. For each study, the effect size of the difference between ASD and TD groups, Hedges' *g*, and its standard error (i.e., the square root of the variance) were calculated based on the mean and standard deviation of the dependent measure. If the information was only available in graphs, it would be extracted using WebPlotDigitizer (Rohatgi, 2019), which is a highly reliable and valid tool to extract graphed data (Drevon et al., 2017). If information on the mean and standard deviation was unavailable, the effect size and its standard error would be estimated using inferential statistics (Thalheimer and Cook, 2002). No study authors were contacted because the effect size could be estimated via one of the aforementioned methods for most studies.

Some studies reported only the mean and standard deviation of the dependent measures for individual emotions but not for overall performance. For these studies, Hedges' *g* and its standard error of group differences in overall performance were estimated using the pooled mean and standard deviation of the dependent measures across emotions[1] . Based on the literature on relationships between basic emotions, r_{ij} was set at 0.27 (Lau et al., 2009). In addition, for studies examining within-subject effects (e.g., static versus dynamic faces), the mean and standard error of the effect size difference were estimated[2] . Since *r* was unknown in all cases, two sets of meta-analyses were performed, with one assuming a zero correlation and the other assuming a large correlation (i.e., $r = 0.5$). Results based on zero correlation are reported because this approach is generally more conservative (i.e., greater variance) than the latter approach; however, the two approaches yielded similar results in all analyses.

For each meta-analysis, the true heterogeneity in effect size across

[1] The pooled standard deviation was calculated using the formula, $\sqrt{var(\frac{1}{m}\sum_{i=1}^{m} Y_i)} = \sqrt{\left(\frac{1}{m}\right)^2(\sum_{i=1}^{m} V_i + \sum_{i\neq j}(r_{ij}\sqrt{V_i}\sqrt{V_j}))}$, where *Y* and *V* denote the dependent measure and the variance of *Y*, respectively, *i* and *j* indicate the i^{th} and j^{th} of the variables (for $i = 1, \ldots, m$), and r_{ij} refers to the correlation between *i* and *j* (Borenstein et al., 2011).

[2] The pooled standard error of the effect size difference was calculated using the formula, $\sqrt{V_{Y_1-Y_2}} = \sqrt{V_{Y_1} + V_{Y_2} - 2r\sqrt{V_{Y_1}}\sqrt{V_{Y_2}}}$, where *V* denotes the variance, Y_1 and Y_2 indicate the effect sizes of two dependent measures, and *r* refers to the correlation between Y_1 and Y_2 (Borenstein et al., 2011).

studies was assessed using the *Q* test (Cochran, 1954). The I^2 statistics describing the percentage of variation across studies that is due to heterogeneity rather than chance are also reported (Higgins and Thompson, 2002). Small, medium and large heterogeneity was indicated by I^2 values of 25 %, 50 %, and 75 %, respectively. For overall performance, meta-regression was performed to explore sources of heterogeneity, including the mean age, sex distribution (i.e., proportion of males), mean IQ (in the following priority order: full-scale > verbal > performance IQ), and symptom severity (i.e., proportion of "autistic disorder" diagnoses, the most severe form of ASD) of the ASD group. A subgroup analysis of studies that involved ASD with comorbidity vs. those that included ASD without any comorbidity was also conducted. The meta-analyses were performed using Meta-Analysis with *R* (Schwarzer et al., 2015).

2.4. Risk of bias and quality assessments

The risk of bias within studies and the quality of each study were assessed by surveying whether ASD and TD groups were matched in age, sex, and IQ. If not, then whether the dependent measure was associated with the matching variable(s) or adjusted for the matching variable(s) was assessed. Additionally, the quality of within-subject design studies was assessed by examining whether the tasks used to measure different levels of a within-subject factor were similar in task type, stimulus

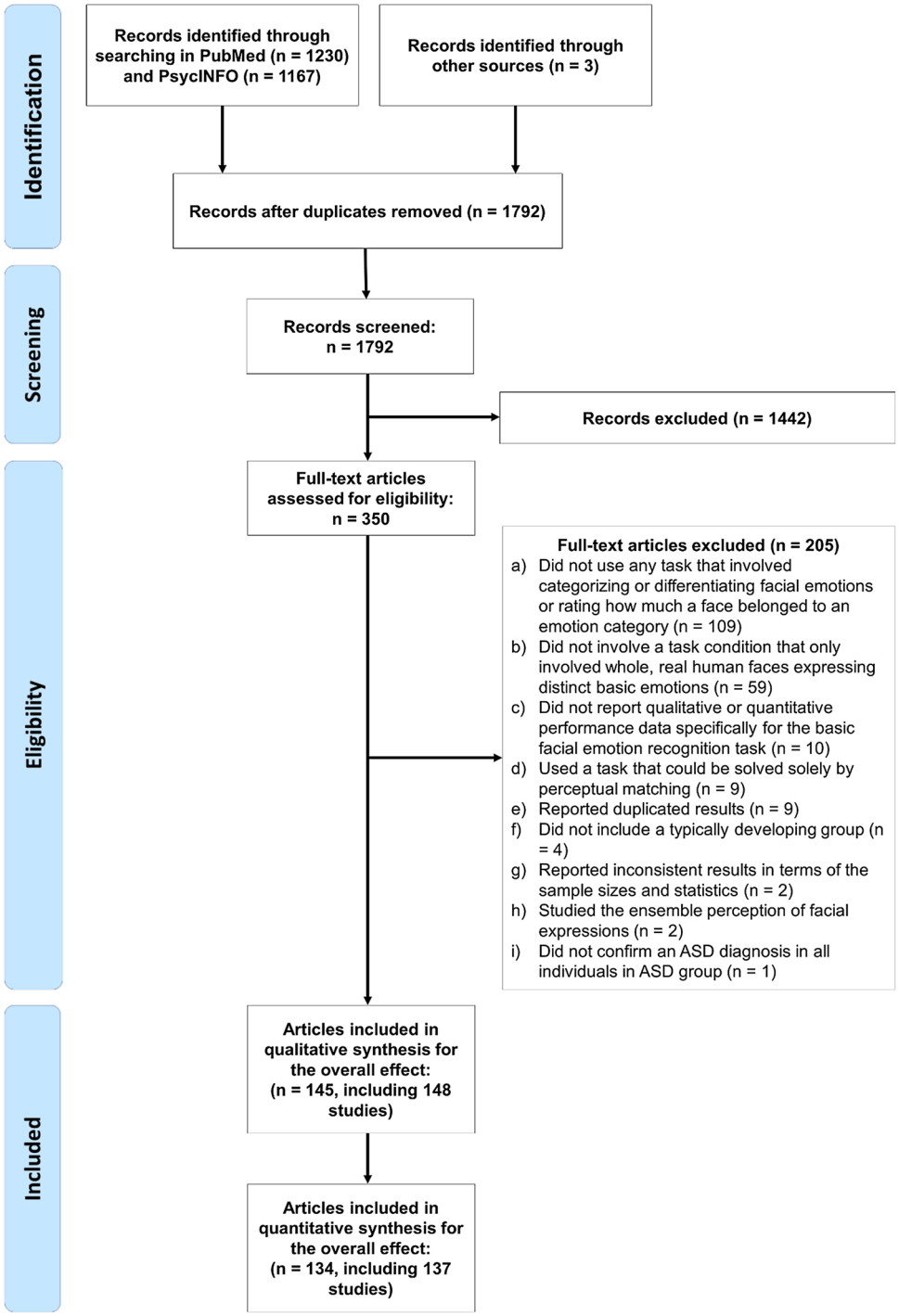


Fig. 1. Flow of the literature search and study selection process.

database, emotion type, and dependent measure. Furthermore, the risk of bias between studies was assessed using funnel plotting and Egger's regression (Egger et al., 1997). The Egger's test was used to evaluate the asymmetry of funnel plots of the effect size of each study and its precision. If a test result was significant, which suggested substantial funnel plot asymmetry (i.e., potential publication bias), Duval and Tweedie's trim-and-fill procedure would be applied to estimate the actual effect size when the "missing" small studies had been published, or when publication bias had been corrected for (Duval and Tweedie, 2000).

A quality assessment based on the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Case Control Studies was also performed (Moola et al., 2020). The checklist consists of ten items; Item 9, which concerns the meaningfulness of the length of the exposure period of interest (i.e., ASD), was not applicable because ASD is a lifelong condition (American Psychiatric Association, 2013; see Appendix B for a complete list of assessment criteria).

3. Results

3.1. Study characteristics

Fig. 1 presents the flow of the literature search. After screening and assessing the eligibility of the identified articles, 205 full-text articles were excluded for the following reasons: a) No task involved categorizing or differentiating facial emotions or rating how much a face belonged to an emotion category (k = 109); b) No task condition involved only full, real human faces expressing distinct basic emotions (k = 59); c) No performance data were reported specifically for the emotion recognition task with basic emotions (k = 10); d) All tasks could be solved by perceptual matching (i.e., without a need to process emotions; k = 9); e) Duplicated results were reported (k = 9); f) No TD group was included (k = 4); g) Inconsistent results were reported (k = 2); h) Ensemble perception was studied (k = 2); and i) Not all individuals in the ASD group had an ASD diagnosis (k = 1).

Thus, 145 articles reporting 148 studies were included in the review. All studies provided enough information for a qualitative or quantitative assessment of facial emotion recognition ability in ASD. Details of these individual studies are presented in Supplementary Table 1. Among these studies, 146 involved an investigation of overall facial emotion recognition ability in ASD. Additionally, 79 studies involved an examination of the effect of basic emotion type on facial emotion recognition ability in ASD. Furthermore, 27 studies compared ASD with other clinical populations, allowing an assessment of the specificity of the facial emotion recognition impairment with respect to the ASD diagnosis. Moreover, 31 and 24 studies adopted a within-subject design and involved an examination of the specificity of such impairment concerning the recognition of emotional facial attributes and to emotion recognition in the face modality in ASD, respectively. Finally, 16, 14, 12, 12, 12, and nine studies have adopted a within-subject design and investigated the impact of stimulus salience, holistic processing, social relevance, emotion complexity, task type, and motion on facial emotion recognition in ASD, respectively.

3.2. Overall facial emotion recognition ability in ASD

All but two of the studies included (i.e., 146) involved a qualitative assessment of overall facial emotion recognition ability in ASD (Supplementary Table 1). Two studies evaluated specific aspects but not the overall level of facial emotion recognition ability in ASD. Impairment in overall facial emotion recognition in ASD was found in 53 % (77/146) of the studies. Among the 146 studies, effect size calculation was not possible for nine studies, eight of which reported null results.

A meta-analysis of overall facial emotion recognition in ASD was performed on 137 studies involving a combined total of 3872 ASD and 4044 TD individuals. A total of 102 studies used a labeling task, 17 studies used a matching task, seven studies used a discrimination task, five studies used a rating task, four studies used a matching and judgment task, one study used a judgment task, and one study used a task that was not described in text. Additionally, static stimuli were used in 124 studies, dynamic stimuli in 12 studies, and a combination of static and dynamic stimuli in one study. Fig. 1 presents a forest plot of the random-effects meta-analysis for overall facial emotion recognition ability in ASD. Autistic individuals had significantly poorer overall facial emotion recognition performance than TD controls, $g = 0.68$, $p < 0.001$, 95 % CI [0.58, 0.78]. A similar pooled effect size was observed when only studies that matched ASD and TD groups on IQ or used an IQ-adjusted dependent measure were included (k = 114), $g = 0.65$, $p < 0.001$, 95 % CI [0.54, 0.76].

There was significant true heterogeneity in effect size across the 137 studies, I^2 = 66 %, $p < 0.001$. However, meta-regression revealed no significant effects of mean age [k = 136; $F(1, 134) = 2.87$, $p = 0.093$], male proportion [k = 131; $F(1, 129) = 0.31$, $p = 0.58$], mean IQ [k = 122; $F(1, 120) = 0.02$, $p = 0.89$], and proportion of "autistic disorder" diagnoses [k = 61; $F(1, 59) = 1.17$, $p = 0.28$] on overall facial emotion recognition ability in ASD. Subgroup analysis also showed no significant effect of comorbidity [k = 26 (15 with comorbidity, 11 without comorbidity); $Q(1) = 1.95$, $p = 0.16$]. Data that formed the basis of these analyses can be found in Supplementary Table 2.

Publication bias was assessed using all 137 studies. Fig. 2a presents a funnel plot of the relationship between the observed effect size and its precision. Egger's test revealed a significant funnel plot asymmetry, $t = 3.66$, $p < 0.001$, suggesting potential publication bias. Therefore, the trim-and-fill procedure was performed to correct for the asymmetry. Fig. 2b illustrates the funnel plot after applying this procedure. After adding 28 studies (k = 165), the pooled effect size (g) was reduced to 0.49 but remained significant, $p < 0.001$, 95 % CI [0.37, 0.61] (Fig. 3).

3.3. Role of basic emotion type in facial emotion recognition ability in ASD

Next, the ability to recognize individual basic emotions in ASD were examined. After excluding the eight studies reporting only a nonsignificant group (ASD, TD) × emotion interaction, and one study reporting significant impairment in the happiness, sadness, anger, and fear combined, and in the disgust and surprise combined, a total of 70 studies were available for a qualitative assessment of these abilities. However, not all basic emotions were examined in each of these studies. Details of these individual studies are presented in Supplementary Table 3. After Bonferroni correction, significant impairment in recognizing the neutral expression was found in 25 % (5/20) of the relevant studies, with the same being found for fear in 21 % (13/61), disgust in 20 % (8/41), sadness in 17 % (10/58), anger in 15 % (10/67), happiness in 8 % (5/64), and surprise in 2 % (1/42) of the studies. Among the 70 studies, effect size calculation was not possible for two studies that reported nonsignificant results for 67–100 % of the emotions. One more study was also excluded from subsequent meta-analysis to avoid bias because effect size calculation was possible for some but not all emotions included in this study.

To determine the ability to recognize individual basic emotions in ASD, separate meta-analyses were performed on studies investigating anger (k = 64), happiness (k = 61), fear (k = 58), sadness (k = 56), disgust (k = 40), surprise (k = 39), and the neutral expression (k = 20). The alpha level was Bonferroni-adjusted to 0.007. Fig. 4a presents a summary of the random-effects meta-analyses for all emotions. The results showed that ASD had significant impairment in recognizing all six basic emotions and the neutral expression, gs from 0.23 to 0.44, ps < 0.001. There was significant heterogeneity in effect size across studies for sadness, anger, fear, and disgust, ps < 0.002, but not for other expressions, ps > 0.013. Egger's test yielded nonsignificant results for all emotions, ps > 0.14.

Thereafter, meta-analyses comparing happiness and other emotions were conducted to determine whether ASD is less impaired in

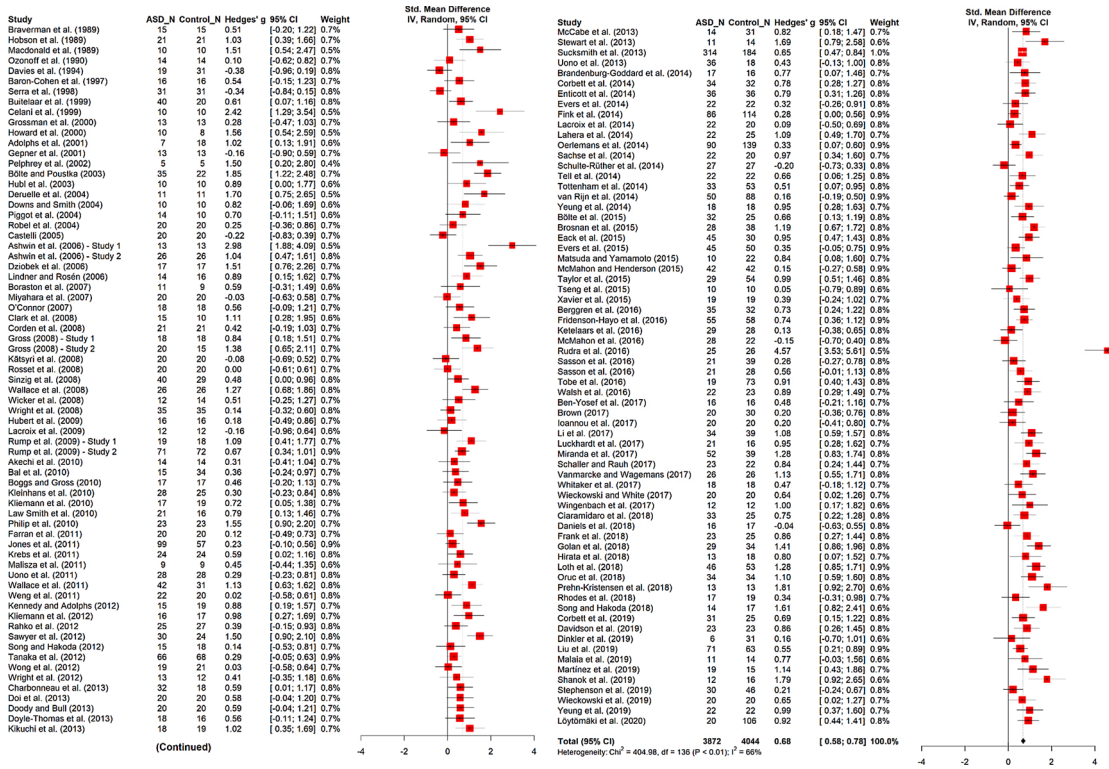


Fig. 2. Forest plot of the random-effects meta-analysis for overall facial emotion recognition ability in autism spectrum disorder compared to typically developing controls.

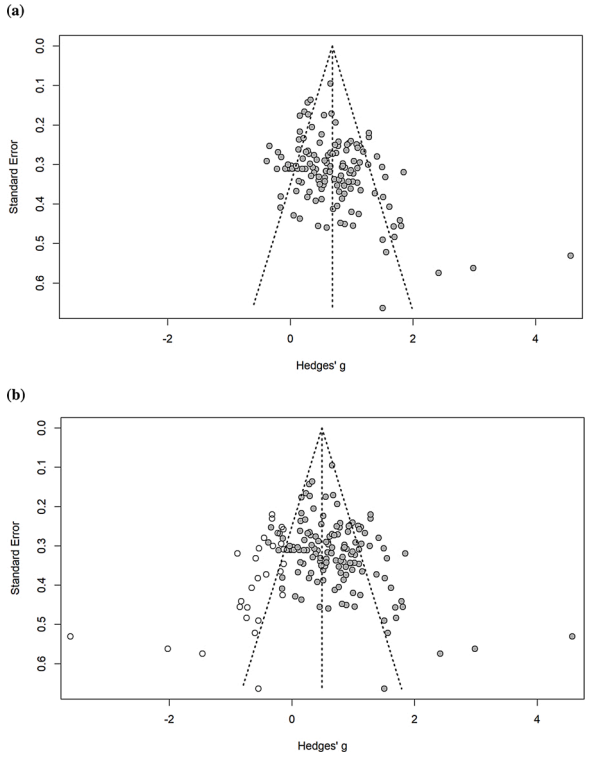


Fig. 3. Funnel plots of the observed effect size and its precision for overall facial emotion recognition in autism spectrum disorder (a) before and (b) after Duval and Tweedie's trim-and-fill procedure. Unfilled circles represent the studies added by this procedure.

recognizing happiness than other emotions, especially negative emotions. The alpha level was Bonferroni-adjusted to 0.008. Fig. 4b presents a summary of the random-effects meta-analyses for the six contrasts. The results indicated that autistic individuals were similarly impaired in recognizing happiness and any other emotion, although there was a statistical trend for these individuals to have greater impairment in recognizing disgust than happiness, *ps* > 0.039. There was only significant heterogeneity in effect size across studies for the fear versus happiness contrast, $I^2 = 37$ %, $p = 0.004$.

3.4. *Specificity of facial emotion recognition impairment in ASD*

3.4.1. *The ASD diagnosis*

Next, the specificity of the ASD diagnosis in facial emotion recognition impairment in ASD was examined. Table 1 presents a summary of this and subsequent sections. A total of 27 studies involved an assessment of whether ASD has poorer facial emotion recognition than other clinical conditions. Three of the studies included two non-ASD groups. The non-ASD clinical conditions included schizophrenia ($k = 7$), ADHD ($k = 6$), language disorder ($k = 4$), mental retardation ($k = 3$), Klinefelter syndrome ($k = 2$), anxiety ($k = 2$), a mixture of ADHD, conduct disorder, and dysthymia ($k = 1$), dyslexia ($k = 1$), conduct disorder ($k = 1$), Down syndrome ($k = 1$), Williams syndrome ($k = 1$), and 22q11.2 deletion syndrome ($k = 1$). Details of these studies are presented in Supplementary Table 4. The ASD versus non-ASD contrasts were significant in 20 % (6/30) of the studies, with ASD having poorer facial emotion recognition than other clinical conditions in all cases.

Except for the two studies (Williams syndrome: $k = 1$; conduct disorder: $k = 1$) reporting a nonsignificant main effect of group (ASD, non-ASD, TD), 25 studies have allowed effect size calculation. Thus, a meta-analysis comparing ASD and non-ASD groups was performed on these 25 studies. For the three studies with more than one non-ASD group, the non-ASD groups were combined into one group. Fig. 5 presents the forest plot of the random-effects meta-analysis for the ASD versus non-ASD contrast. The meta-analysis showed that ASD had significantly

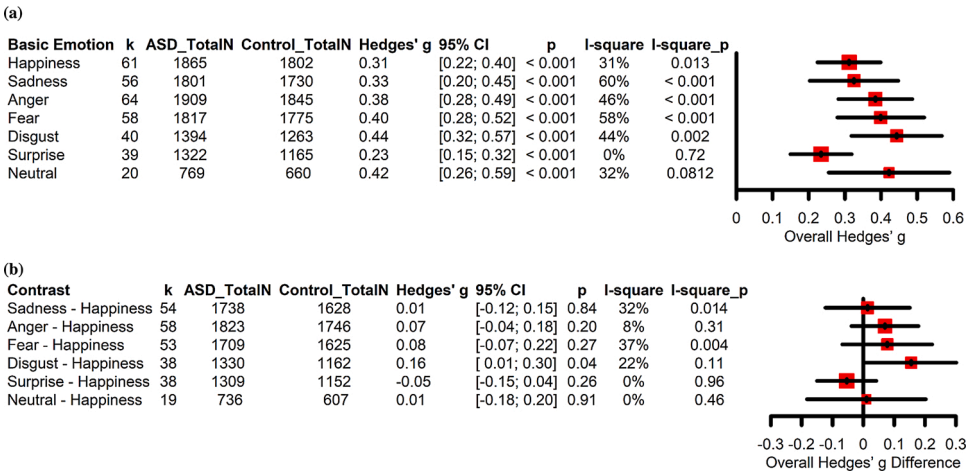


Fig. 4. Summary of the random-effects meta-analyses (a) for individual emotions and (b) for the contrasts between happiness and other emotions in autism spectrum disorder.

poorer facial emotion recognition than other clinical conditions, $g = 0.26$, $p = 0.017$, 95 % CI [0.05, 0.48]. Heterogeneity in effect size across studies was significant, $I^2 = 62$ %, $p < 0.001$.

3.4.2. Recognition of emotional facial attributes

To determine whether ASD has poorer recognition of emotional than nonemotional facial attributes, 31 studies that have allowed an assessment of this comparison were examined. Details of these studies are presented in Supplementary Table 5. All but two of these studies used one nonemotional task involving the recognition of identity ($k = 15$), sex ($k = 10$), age ($k = 3$), both identity and sex ($k = 1$), and facial motion ($k = 1$). Two studies used two nonemotional tasks, with one study using identity and sex recognition tasks and another study using identity and gaze direction recognition tasks. The identity task was chosen over the gaze direction or sex task to represent the nonemotion task for subsequent analysis. The emotion and nonemotion tasks were similar in task type, stimulus database, emotion type, and dependent measure in 61 % (19/31) of the studies. Overall, one study reported a nonsignificant interaction between group (ASD, TD) and condition (emotion, sex). For the remaining 30 studies, ASD was impaired on both tasks in seven studies, on neither task in 10 studies, only on the emotion task in 10 studies, and only on the nonemotion task in three studies.

A meta-analysis comparing facial emotion and nonemotion tasks was performed on 29 studies, excluding the study that only reported a nonsignificant interaction effect and another study that allowed effect size calculation only for the emotion task. The result showed that the effect sizes of group difference between the emotion and nonemotion tasks (emotion > nonemotion) were not significantly different, $g = 0.12$, $p = 0.36$, 95 % CI [−0.15, 0.40]. Heterogeneity in effect size across studies was significant, $I^2 = 47$ %, $p = 0.003$.

3.4.3. Emotion recognition in face modality

To determine whether ASD is more impaired in recognizing emotion in faces than in other modalities, 24 studies that have allowed an assessment of this contrast were analyzed. Details of these studies are presented in Supplementary Table 6. All but five studies used one nonface task involving the recognition of emotion in voices ($k = 14$), bodies ($k = 2$), scenes ($k = 2$), and verbal content ($k = 1$). Moreover, five studies used two nonface tasks, with two studies using a voice and a body emotion task, one study using a voice and a verbal content emotion task, one study using a scene and a point light emotion task, and one study using a posture and a gesture emotion task. The voice, scene, and posture tasks were chosen over the body or verbal content, point light, and gesture tasks for subsequent analysis. The face and nonface tasks were similar in task type, emotion type, and dependent measure in 75 % (18/24) of the studies. Overall, one study reported (marginally) significant differences between the ASD and TD groups on both the face and nonface tasks. For the remaining 23 studies, ASD was impaired on both tasks in six studies, on neither task in 12 studies, only on the face task in four studies, and only on the nonface task in one study.

A meta-analysis comparing the ability to recognize emotion in faces and other modalities was performed on 23 studies, excluding the study reporting (marginally) significant impairment of unknown effect size on both the face and nonface tasks in ASD. The result showed that the effect sizes of group difference between the face and nonface tasks (face > nonface) were not significantly different, $g = 0.09$, $p = 0.21$, 95 % CI [−0.06, 0.24]. There was no significant heterogeneity in effect size across studies, $I^2 = 0$ %, $p = 0.75$.

3.5. Impact of task characteristics on facial emotion recognition in ASD

3.5.1. Task type

To determine whether ASD has poorer performance on tasks that do not involve language than on tasks with verbal cues, 12 studies that have allowed an assessment of this contrast were analyzed. Details of these studies are presented in Supplementary Table 7. A total of eight studies used a labeling and a matching task, while three studies used a labeling, a matching, and a discrimination task, and one study used a discrimination and a matching task. For studies using two verbal tasks, the labeling task was chosen over the discrimination task for subsequent analysis. The verbal and nonverbal tasks were similar in task type, stimulus database, emotion type, and dependent measure in 92 % (11/12) of the studies. Overall, ASD was impaired on both tasks in two studies, on neither task in eight studies, and only on the nonverbal task in two studies. No studies found impairment on only the verbal task.

A meta-analysis comparing the verbal and nonverbal tasks was performed on 10 studies that have allowed effect size calculation. The result identified a statistical trend for ASD to have poorer performance on nonverbal compared to verbal tasks, $g = 0.27$, $p = 0.075$, 95 % CI [−0.04, 0.58]. True heterogeneity in effect size across studies was not significant, $I^2 = 6\%$, $p = 0.38$.

3.5.2. Motion

To determine whether performance on tasks with static and dynamic stimuli was different in ASD, nine studies that have allowed an assessment of this contrast were analyzed. Details of these studies are presented in Supplementary Table 8. Tasks with static and dynamic stimuli involved the same type of task: labeling in seven studies, matching in one study, and rating in one study. Additionally, the static and dynamic tasks were similar in task type, stimulus database, emotion type, and

Table 1
Summary of the assessment of the specificity of facial emotion recognition impairment and the role of task characteristics in facial emotion recognition in autism spectrum disorder (ASD).

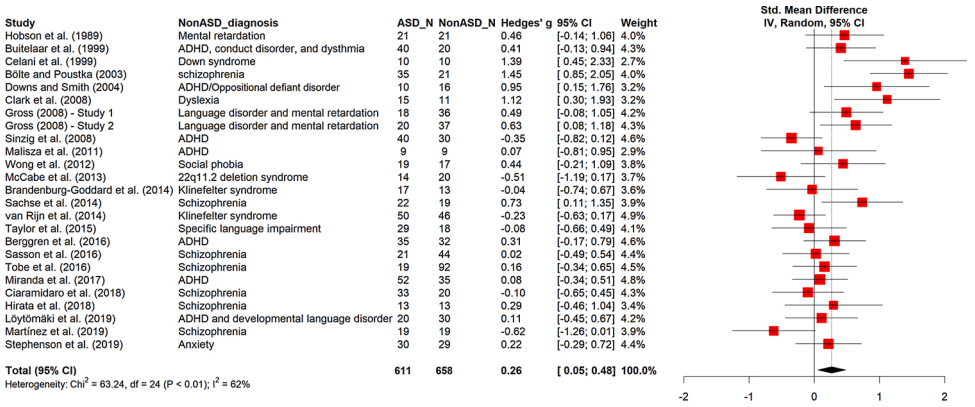
Factor	Number of studies (*k*)	Condition 1	Condition 2	Matching in task type, stimulus database, emotion type, and measure (%)	Impairment in both conditions, neither condition, Condition 1 only, Condition 2 only, and unknown conditions (*k*)	Number of studies in meta-analysis	Total ASD subjects per condition	Total TD subjects per condition	Hedges' *g* difference (Condition 1 > Condition 2) [95% CI]	*p*
(a) Specificity of impairment										
Diagnosis	27	ASD	Non-ASD clinical conditions	/	8, 10, 7, 1, 1	25 (93 %)	611	658	0.26[#] [0.05, 0.48]	0.017 *
Emotion	31	Emotional facial attributes	Nonemotional facial attributes	71, 68, 61, 94	7, 10, 10, 3, 1	29 (94 %)	602–605	578–584	0.12 [-0.15, 0.40]	0.36
Modality	24	Facial emotions	Vocal and body emotions	92, /, 79, 96	6, 12, 4, 1, 1	23 (96 %)	645–667	689–756	0.09 [-0.06, 0.24]	0.21
(b) Role of task characteristics										
Task type	12	Tasks with verbal cues	Tasks without verbal cues	/, 92, 100, 100	2, 8, 0, 2, 0	9 (75 %)	237–239	239–241	−0.27 [-0.58, 0.04]	0.075[†]
Motion	9	Static faces	Dynamic faces	100, 78, 89, 100	1, 4, 2, 1, 1	8 (89 %)	179	278	0.07 [-0.24, 0.38]	0.60
Emotion complexity	12	Basic emotions	Complex emotions	92, 50, 67, 92	7, 2, 1, 1, 1	10 (83 %)	284	275	−0.26 [-0.45, -0.06]	0.016 *
Holistic processing	15	Holistic processing	Feature-based processing	100, 100, 100, 100	6, 6, 1, 1, 1	14 (93 %)	318	292	0.31 [0.03 0.58]	0.031 *
Social relevance	12	High social relevance	Low social relevance	100, 67, 83, 100	0, 7, 3, 1, 1	11 (92 %)	207	221	0.26 [-0.24, 0.76]	0.27
Stimulus salience	16	Low salience	High salience	100, 100, 100, 100	0, 6, 3, 2, 5	11 (69 %)	192	214	−0.30 [-0.71 0.12]	0.14

*Note.*TD = typically developing. Significant impairment was determined by the *p*-values of statistical tests comparing ASD and typically developing groups after Bonferroni correction ($p < 0.05$). Except for diagnosis, random-effects meta-analysis was performed on the effect size difference between two conditions, assuming a zero correlation between the two effect sizes.

[#] Estimated based on the effect size of contrast between ASD and non-ASD clinical groups.

[†] $p < 0.10$.

[*] $p < 0.05$.



Study	NonASD_diagnosis	ASD_N	NonASD_N	Hedges' g	95% CI	Weight
Hobson et al. (1989)	Mental retardation	21	21	0.46	[-0.14; 1.06]	4.0%
Buitelaar et al. (1999)	ADHD, conduct disorder, and dysthmia	40	20	0.41	[-0.13; 0.94]	4.3%
Celani et al. (1999)	Down syndrome	10	10	1.39	[0.45; 2.33]	2.7%
Bölte and Poustka (2003)	schizophrenia	35	21	1.45	[0.85; 2.05]	4.0%
Downs and Smith (2004)	ADHD/Oppositional defiant disorder	10	16	0.95	[0.15; 1.76]	3.2%
Clark et al. (2008)	Dyslexia	15	11	1.12	[0.30; 1.93]	3.2%
Gross (2008) - Study 1	Language disorder and mental retardation	18	36	0.49	[-0.08; 1.05]	4.2%
Gross (2008) - Study 2	Language disorder and mental retardation	20	37	0.63	[0.08; 1.18]	4.3%
Sinzig et al. (2008)	ADHD	40	30	-0.35	[-0.82; 0.12]	4.6%
Malisza et al. (2011)	ADHD	9	9	0.07	[-0.81; 0.95]	2.9%
Wong et al. (2012)	Social phobia	19	17	0.44	[-0.21; 1.09]	3.8%
McCabe et al. (2013)	22q11.2 deletion syndrome	14	20	-0.51	[-1.19; 0.17]	3.7%
Brandenburg-Goddard et al. (2014)	Klinefelter syndrome	17	13	-0.04	[-0.74; 0.67]	3.6%
Sachse et al. (2014)	Schizophrenia	22	19	0.73	[0.11; 1.35]	3.9%
van Rijn et al. (2014)	Klinefelter syndrome	50	46	-0.23	[-0.63; 0.17]	4.9%
Taylor et al. (2015)	Specific language impairment	29	18	-0.08	[-0.66; 0.49]	4.1%
Berggren et al. (2016)	ADHD	35	32	0.31	[-0.17; 0.79]	4.6%
Sasson et al. (2016)	Schizophrenia	21	44	0.02	[-0.49; 0.54]	4.4%
Tobe et al. (2016)	Schizophrenia	19	92	0.16	[-0.34; 0.65]	4.5%
Miranda et al. (2017)	ADHD	52	35	0.08	[-0.34; 0.51]	4.8%
Ciaramidaro et al. (2018)	Schizophrenia	33	20	-0.10	[-0.65; 0.45]	4.3%
Hirata et al. (2018)	Schizophrenia	13	13	0.29	[-0.46; 1.04]	3.4%
Löytömäki et al. (2019)	ADHD and developmental language disorder	20	30	0.11	[-0.45; 0.67]	4.2%
Martínez et al. (2019)	Schizophrenia	19	19	-0.62	[-1.26; 0.01]	3.9%
Stephenson et al. (2019)	Anxiety	30	29	0.22	[-0.29; 0.72]	4.4%
Total (95% CI)		**611**	**658**	**0.26**	**[0.05; 0.48]**	**100.0%**

Heterogeneity: $Chi^2 = 63.24$, df = 24 (P < 0.01); $I^2 = 62\%$

Fig. 5. Forest plot of the random-effects meta-analysis for overall facial emotion recognition ability in autism spectrum disorder compared to other clinical conditions. ADHD = attention-deficit/hyperactivity disorder.

dependent measure in 78 % (7/9) of the studies. Overall, one study reported a nonsignificant main effect of group (TD, ASD) and a nonsignificant interaction between group and motion (static, dynamic). For the remaining studies, ASD was impaired on both tasks in one study, on neither task on four studies, on only the static task in two studies, and on only the dynamic task in one study.

A meta-analysis comparing performance on static and dynamic tasks was performed on eight studies that have allowed effect size calculation. The result showed that the effect sizes of group difference between the static and dynamic tasks (static > dynamic) were not significantly different, $g = 0.07$, $p = 0.60$, 95 % CI [−0.24, 0.38]. Heterogeneity in effect size across studies was not significant, $I^2 = 0\%$, $p = 0.56$.

3.5.3. Emotion complexity

To determine whether ASD is poorer at recognizing complex than basic facial emotions, 12 studies that have allowed an assessment of this contrast were analyzed. Details of these studies are presented in Supplementary Table 9. Basic emotion tasks involved three basic emotions in one study, four basic emotions in two studies, and all six basic emotions in nine studies. Complex emotion tasks involved a variety of emotions, such as friendly, playful, interested, etc. The Reading the Mind in the Eyes task, which requires the forced choice labeling of complex emotions (Baron-Cohen et al., 1997), was used in half of these studies. The basic and complex tasks were similar in task type, stimulus database, emotion type, and dependent measure in 50 % (6/12) of the studies. Overall, one study reported a significant impairment on the complex task and marginally significant or significant impairment on the basic task in ASD. In the remaining studies, ASD was impaired on both tasks in seven studies, on neither task in two studies, on only the basic task in one study, and on only the complex task in one study.

A meta-analysis comparing the basic and complex tasks was performed on 10 studies that have allowed effect size calculation. The result showed that ASD had significantly poorer performance on the complex than the basic emotion tasks, $g = 0.26$, $p = 0.016$, 95 % CI [0.06, 0.45]. Heterogeneity in effect size across studies was not significant, $I^2 = 0\ \%$, $p = 0.93$.

3.5.4. Holistic processing

To determine whether ASD has poorer performance on facial emotion recognition tasks that emphasize holistic more than featural face processing, 15 studies that have allowed an assessment of this contrast were analyzed. Details of these studies are presented in Supplementary Table 10. The orientation effect was examined in six studies, the face integrity effect was examined in six studies, and the spatial frequency effect was examined in three studies. Notably, the strong low-pass filtering condition was chosen over the slight low-pass filtering condition to represent the holistic task in one study. Additionally, one study investigated both the orientation and spatial frequency effects. To avoid including the same study twice, the condition with upright and low-pass filtered faces was used to represent the holistic task, while the condition with inverted and high-pass filtered faces ($g = 0.25$, $SE = 0.32$) was used to represent the featural task. Finally, the hybridity effect was investigated in one study, with a neutral mouth chosen over neutral eyes to represent the featural task.

The holistic (i.e., upright, whole-face, low-pass filtered, or nonhybrid facial stimuli) and featural (i.e., inverted or rotated, eyes-only, high-pass filtered, or hybrid facial stimuli) tasks were similar in task type, stimulus database, emotion type, and dependent measure in all 15 studies. Overall, one study on the orientation effect reported a nonsignificant main effect of group (TD, ASD) and a nonsignificant interaction between group and orientation (30, 90, 150, 180, 210, 270, 330 rotation degrees). For the others, ASD was impaired on both tasks in six studies, on neither task in six studies, on only the holistic task in one study, and on only the featural task in one study.

A meta-analysis comparing the holistic and featural tasks was performed on 14 studies that have allowed effect size calculation. The result showed that ASD had significantly poorer performance on the holistic than the featural tasks, $g = 0.31$, $p = 0.031$, 95 % CI [0.03, 0.58]. There was no significant heterogeneity in effect size across studies, $I^2 = 15\ \%$, $p = 0.29$.

3.5.5. Social relevance

To determine whether ASD is poorer at recognizing emotions from more socially relevant than less socially relevant faces, 12 studies that have allowed an assessment of this contrast were analyzed. Details of these studies are presented in Supplementary Table 11. The effect of humanity was examined in six studies. One study included a human cartoon and a nonhuman cartoon condition, and the nonhuman cartoon condition was chosen to represent the socially irrelevant task. Additionally, the effect of gaze direction was examined in four studies, the effect of ethnicity was investigated in one study, and the effect of familiarity was examined in one study. The socially relevant (i.e., familiar, real human, same-ethnicity, or direct-gaze faces) and irrelevant (i.e., unfamiliar, nonhuman, different-ethnicity, or averted-gaze faces) tasks were similar in task type, stimulus database, emotion type, and dependent measure in 58 % (7/12) of the studies. Overall, one study reported a significant main effect of group (TD, ASD) and a nonsignificant interaction between group and gaze direction (direct, averted). For the others, ASD was impaired on neither task in seven studies, on only the socially relevant task in three studies, and on only the socially irrelevant task in one study. None of the studies showed impairment on both tasks.

A meta-analysis comparing the socially relevant and irrelevant tasks was performed on 11 studies that have allowed effect size calculation. The result showed that the effect sizes of group difference between the

socially relevant and irrelevant tasks (relevant > irrelevant) were not significantly different, $g = 0.26$, $p = 0.27$, 95 % CI [−0.24, 0.76]. Heterogeneity in effect size across studies was significant, $I^2 = 64$ %, $p = 0.002$.

3.5.6. Stimulus salience

Finally, the role of stimulus salience in terms of presentation time ($k = 3$) and emotion intensity ($k = 13$) was examined. Details of these studies are presented in Supplementary Table 12. In all studies, the same task type, stimulus database, emotion type, and dependent measure were used for different factor levels. A parametric design with more than two factor levels was adopted in 77 % (10/13) of the studies. The relationship between the level of stimulus salience and the effect size of group difference was flat in three studies, an inverted U-shape in three studies, a monotonic increase in two studies, a U-shape in one study, and unknown in one study. Overall, five studies only reported a nonsignificant interaction between group (ASD, TD) and presentation time ($k = 1$) or emotion intensity ($k = 4$). Among these five studies, one study on emotion intensity showed a significant group effect, whereas others observed a nonsignificant group effect. For the remaining studies, the conditions with the highest salience (i.e., longest presentation time or strongest emotion intensity) and lowest salience (i.e., shortest presentation time or weakest emotion intensity) were chosen for subsequent analysis. No studies observed impairment in both conditions in ASD, and ASD was found to be impaired in neither condition in six studies, on only the lowest salience condition in three studies, and in only the highest salience condition in two studies.

A meta-analysis comparing the highest and lowest salience conditions was performed on 11 studies, excluding the five studies reporting a nonsignificant interaction effect. The result showed that the effect sizes of group difference between the two salience conditions (highest > lowest) were not significantly different, $g = 0.30$, $p = 0.14$, 95 % CI [−0.12, 0.71]. There was no significant heterogeneity in effect size across studies, $I^2 = 39$ %, $p = 0.092$.

Table 2
Overall quality assessment results based on the JBI Critical Appraisal Checklist for Case Control Studies.

	Yes (*k*; %)	Unclear (*k*; %)	No (*k*; %)	N/A (*k*; %)
1. Were the groups comparable other than the presence of disease in cases or the absence of disease in controls?	89 (60 %)	21 (14 %)	38 (26 %)	0 (0 %)
2. Were cases and controls matched appropriately?	55 (37 %)	68 (46 %)	25 (17 %)	0 (0 %)
3. Were the same criteria used for identification of cases and controls?	78 (53 %)	69 (47 %)	1 (1 %)	0 (0 %)
4. Was exposure measured in a standard, valid and reliable way?	131 (89 %)	14 (9 %)	3 (2 %)	0 (0 %)
5. Was exposure measured in the same way for cases and controls?	68 (46 %)	70 (47 %)	10 (7 %)	0 (0 %)
6. Were confounding factors identified?	145 (98 %)	0 (0 %)	3 (2 %)	0 (0 %)
7. Were strategies to deal with confounding factors stated?	128 (86 %)	1 (1 %)	6 (4 %)	13 (9 %)
8. Were outcomes assessed in a standard, valid and reliable way for cases and controls?	137 (93 %)	8 (5 %)	3 (2 %)	0 (0 %)
9. Was the exposure period of interest long enough to be meaningful?	0 (0 %)	0 (0 %)	0 (0 %)	148 (100 %)
10. Was appropriate statistical analysis used?	40 (27 %)	105 (71 %)	3 (2 %)	0 (0 %)

Note. Item 9 was inapplicable for all studies because autism spectrum disorder is a lifelong condition.

3.6. Study quality assessment

Table 2 presents the overall quality assessment results, and Supplementary Table 13 presents the assessment results for each individual study. Among all the 148 studies, the groups were comparable in key demographic features (i.e., age, sex, and IQ) other than the ASD diagnosis in 89 (60 %) studies. Cases and controls were recruited from the same source population and were thus matched appropriately in 55 (37 %) studies; however, this information was either unclear or missing in most other studies. Inclusion/exclusion criteria used for identification of cases and controls were the same in 78 (53 %) studies; for others, there was insufficient information regarding whether the same criteria were applied to both groups. The ASD construct was assessed in a standard, valid and reliable way in 131 (89 %) studies. Such assessment was conducted in the same way for cases and controls in 68 (46 %) studies; for others, whether the ASD evaluation was similarly carried out for both groups was unclear. Confounding factors were identified in 145 (98 %) studies, and strategies to deal with confounding factors (e.g., group matching on IQ) were stated in 128 (86 %) studies. Outcomes (i.e., facial emotion recognition performances) were assessed in a standard, valid and reliable way in 137 (93 %) studies. Statistical analysis was appropriate (e.g., checking of test assumptions) in 40 (27 %) studies, whereas the appropriateness of statistical methodology was unclear for most other studies.

4. Discussion

This study conducted qualitative and quantitative assessments of facial emotion recognition in ASD, assessed the specificity of facial emotion recognition impairment, if any, and the role of task characteristics in facial emotion recognition in ASD. The results of a meta-analysis suggest that ASD has impairment in recognizing basic facial emotions overall and in recognizing all individual emotions. Moreover, ASD has poorer facial emotion recognition than other clinical populations and has comparable impairment in recognizing emotional and nonemotional facial attributes, as well as in recognizing emotion in faces and other modalities. Regarding task characteristics, while there is no evidence that motion, social relevance, and stimulus salience have a moderating effect on facial emotion recognition performance in ASD, emotion complexity and holistic processing do. There is also a trend suggesting that task type could be a moderator of facial emotion recognition performance in ASD.

The present findings are consistent with previous meta-analytic findings of moderate impairment in overall facial emotion recognition in ASD (Lozier et al., 2014; Uljarevic and Hamilton, 2013). Consistent with Uljarevic and Hamilton (2013), but not with Lozier et al. (2014), publication bias was found to be present in the literature; however, it does not have a significant impact on the result. Two sources of publication bias were identified. First, the asymmetrical funnel plot of the observed effect size and its precision suggests that small studies reporting positive results were more likely than small studies reporting null results to be published in peer-reviewed journals. Additionally, studies that provided insufficient information for effect size calculation were likely to have reported null results. These findings have two implications for editors and authors. First, small studies reporting significant effects should be taken seriously during the publication stage. Second, regardless of whether their results are significant or not, future studies should report the test statistics and exact *p*-value, and preferably the mean and standard deviation, to enable an unbiased synthesis of the literature.

It should be emphasized that the present meta-analytic results indicate mean differences between autistic and TD individuals, not universal group differences. There existed large individual differences within ASD and significant true heterogeneity in effect size across studies, but the mean age, sex, IQ, symptom severity, and psychiatric comorbidity of the ASD group failed to significantly explain this heterogeneity. These

inconclusive results may be due to the heterogeneous study design across studies, and are consistent with those reported in Uljarevic and Hamilton (2013), where neither age nor IQ was identified as significant moderators. In contrast, Lozier et al. (2014) found that age (but not IQ) significantly moderated overall facial emotion recognition performance in ASD. The moderator analysis performed by Uljarevic and Hamilton (2013) and the present study considered the sample size of each study, whereas the analysis done by Lozier et al. (2014) did not. Thus, the discrepancy across studies could be due to methodological differences.

One of the major contributions of this review is that it helps to resolve some of the controversial issues regarding differential impairment in basic emotion recognition in ASD. In two previous meta-analyses, Uljarevic and Hamilton (2013) found significant impairment in all basic emotions except happiness, while Lozier et al. (2014) reported significant impairment in happiness, anger, fear, and surprise, but not in others, after multiple testing correction. One strength of the present meta-analysis is that it was based on approximately three times as many studies as the previous meta-analyses, which included studies published before 2012 (Lozier et al., 2014; Uljarevic and Hamilton, 2013). The large number of studies included in the present meta-analysis is possibly due to increased recent research activity since 58 % of the included studies were published since 2012, and to the use of WebPlotDigitizer (first released in around 2011) to extract data from figures. Thus, the present study had more statistical power to assess the ability to recognize individual emotions in ASD. The present findings of significant impairment in recognizing all basic facial emotions—even after Bonferroni correction—suggest nonselective impairment in ASD.

Furthermore, there is a tendency for ASD to be more impaired in recognizing disgust than happiness. Given that the effect size of group difference for disgust was numerically higher than that for all other emotions, it remains to be determined whether disgust is the most poorly recognized basic facial emotion in ASD. Unlike in the meta-analysis of Uljarevic and Hamilton (2013), a trend for ASD to be more impaired in recognizing fear than happiness was not found in the present meta-analysis. Based on an analysis of heterogeneity, this lack of effect might be attributable to the large heterogeneity in effect size across studies, thus implying that the ability to recognize fear relative to happiness in ASD greatly depends on the characteristics of the sample and task included. Additionally, this discrepancy in findings might be due to methodological differences across studies. Specifically, the present comparisons of happiness with other basic emotions were based on 38–56 facial emotion recognition studies, whereas the comparisons by Uljarevic and Hamilton (2013) were based on 10–16 studies on emotion recognition in the visual modality. Moreover, while we used a random-effects meta-analysis and assigned different weights to studies varying in sample size, Uljarevic and Hamilton (2013) used paired *t*-tests, thereby assigning the same weight to all studies.

The present meta-analysis on the contrast between ASD and non-ASD populations suggests that facial emotion recognition is generally poorer in ASD than in other psychiatric and medical conditions that share similar traits and are commonly comorbid with ASD (Matson & Goldin, 2013). Since facial emotion recognition ability is crucial for social competence (Izard et al., 2001) and social communication and interaction impairment is a diagnostic feature of ASD (American Psychiatric Association, 2013), this finding suggests that measures of facial emotion recognition may potentially serve as a marker for ASD. Nevertheless, all non-ASD clinical conditions were combined into a single group because few studies have compared ASD using the same clinical population. Subsequently, the included non-ASD sample was heterogeneous. Notably, more research is required to determine whether ASD has poorer facial emotion recognition than specific clinical populations that are commonly comorbid with ASD, such as ADHD.

Facial emotion recognition involves the visuoperceptual processing of faces followed by the conceptual analysis of the emotion conveyed by the face (Adolphs, 2002). The present meta-analysis showed that ASD has comparable impairment in recognizing emotional and nonemotional facial attributes, thus raising the possibility that deficits in facial perception, such as impaired facial identity recognition (Weigelt et al., 2012), may underlie impaired facial emotion recognition in ASD. Indeed, some evidence suggests that facial perception impairment is associated with deficits in recognizing negative expressions in ASD (Yeung et al., 2019); however, this does not fully explain the facial emotion recognition impairment. Furthermore, the present meta-analysis found that ASD has similar emotion recognition deficits across modalities. This finding may reflect common psychological and neurological mechanisms underlying emotion perception in various communication channels (Schirmer and Adolphs, 2017) and suggests that impairment in linking perceptual representations and emotion concepts—and in retrieving emotional knowledge—may underlie the poor facial emotion recognition ability in ASD. Altogether, some evidence suggests that deficits in both the basic perception of faces and the conceptual analysis of emotion underpin facial emotion recognition impairment in ASD.

One unique contribution of this work is that the role of task characteristics in facial emotion recognition in ASD was systematically assessed by taking advantage of within-subject design studies. Thus, the factors of interest can be assessed with minimal confounding effects from other factors, especially sample characteristics. More importantly, whether the tasks used to measure different levels of each factor of interest were matched in task type, stimulus database, emotion type, and dependent measure were thoroughly surveyed to determine whether the effect of a factor could be possibly confounded by other factors. This assessment revealed that the tasks used were comparable in each of the four aspects in 50–100 % of the studies included in each meta-analysis. Additionally, it was found that 69–96 % of the studies examining one task factor were included in the meta-analysis for that factor. Thus, the meta-analysis results seem representative of the overall effects of the factors of interest. Nonetheless, publication bias, in which null results were not reported even without the provision of effect size information, may limit the amount of evidence available for such analyses.

Emotion complexity, holistic processing, and task type—but not motion, social relevance, or stimulus salience—were identified as potential moderators of facial emotion recognition performance in ASD. Specifically, the greater impairment in recognizing complex compared to basic emotions may reflect deficits in inferring others' mental states in ASD (Baron-Cohen et al., 1997). Alternatively, complex emotions are harder to recognize, and hence more sensitive to impairment (Fridenson-Hayo et al., 2016). Additionally, poorer performance on tasks that emphasize holistic rather than featural face processing may be related to the detail-focused cognitive style (Happé and Frith, 2006) and impaired facial perception in ASD. In fact, experimental studies with healthy people have shown that holistic processing generally contributes more strongly than featural processing to expression recognition (Bombari et al., 2013; Derntl et al., 2009). Thus, failure to employ holistic processing may underpin the facial emotion recognition deficits in ASD. Moreover, poorer performance on matching tasks than on tasks with verbal cues could be related to the impaired retrieval of conceptual knowledge that facilitates facial expression categorization (Lindquist et al., 2006; Nook et al., 2015) in ASD. Despite these findings, the effect of task type should be taken with caution since the result is only marginally significant and must be verified by further studies.

The very small, if not absent, effect of motion on facial emotion recognition in ASD is consistent with the claim that ASD may be able to benefit from the dynamic presentation of facial expressions in general (Enticott et al., 2014; Gepner et al., 2001). It also suggests that tasks using static faces, although lacking ecological validity, remain useful for assessing facial emotion recognition in ASD. Additionally, the nonsignificant effect of social relevance suggests that ASD may have comparable emotion recognition deficits across face types. There was a moderate-to-large heterogeneity in effect size across studies, implying that although it was hypothesized that familiar, human, same-ethnicity, and direct-gaze faces are generally more socially relevant than

unfamiliar, nonhuman, different-ethnicity, and averted-gaze faces, respectively, each of these facial attributes may have a different moderating effect on emotion recognition performance in ASD. To address this issue, additional studies examining the effect of each of these facial attributes are required.

The nonsignificant and inconsistent effect of stimulus salience suggests that ASD may have similar facial emotion recognition deficits across salience levels. Alternatively, this could be due to the use of different stimulus presentation times and emotion intensity levels across studies. Furthermore, the analysis is complicated by several studies including a task condition in which the floor effect might even be present in TD controls (Clark et al., 2008; Doi et al., 2013; Rhodes et al., 2018). Thus, further studies adopting a parametric design and employing tasks that span a full range of salience levels without hitting the floor effect are required to better clarify the impact of salience levels. Some experimental studies with healthy people have found that the accuracies in labeling facial emotions when facial stimuli are presented for 25, 50–100, and $\geq$ 250 ms (Calvo and Lundqvist, 2008), and when the emotion intensity levels are 40–50, 60–70, and 80–100 % of prototypical expressions (Hoffmann et al., 2010), are above chance level and distinguishable across conditions. Therefore, future studies are recommended to employ these parameters to study the salience effect.

Over the last two decades, eye-tracking and electroencephalography studies have mostly reported atypical gaze to core facial features and atypical modulation of the N170 event-related potential component, respectively, during facial emotion recognition tasks in ASD (Black et al., 2017). Although the results have been mixed, functional neuroimaging studies have also reported altered activity in the amygdala, fusiform face area, superior temporal sulcus, and inferior frontal gyrus during emotional face processing in ASD (Harms et al., 2010; Philip et al., 2012). These altered face scanning and activation in face-selective, emotion, and mentalizing brain regions may underpin some of the effects observed in the present study (e.g. similar impairment on facial emotion and nonemotion tasks and relatively greater impairment on tasks emphasizing complex emotion recognition and holistic face processing in ASD). Notably, some studies have found that motion, familiarity, and emotion intensity modulate brain responses in TD but not ASD individuals (Philip et al., 2012). The lack of these effects on facial emotion recognition ability in ASD could be due to the utilization of compensatory strategies (e.g., featural processing) during task performance.

The quality assessment of studies revealed some limitations in the existing literature. Specifically, many studies have not clearly reported information about the source population and the inclusion/exclusion criteria used to identify autistic and TD individuals. Such information should be reported with sufficient detail because it sheds light on whether sampling bias is present. In addition, for many studies, it is often unclear whether a valid and reliable ASD or psychiatric evaluation was carried out for TD individuals. Conducting such evaluation using standardized instruments is important for ensuring that TD individuals do not exhibit a clinical level of autistic symptoms or meet the criteria for any psychiatric disorder. Furthermore, many studies appear to have either neglected statistical test assumptions or failed to report such information before comparing groups. The assumptions of statistical tests (e.g., the normality assumption and the homogeneity of variance assumption of *t*-tests and ANOVA) should be checked to make sure that the obtained results are valid.

This study also has some limitations. First, only two databases (PubMed and PsycINFO) were used to identify studies, which may limit the exhaustiveness of the search. Second, while a second individual was involved in checking the study selection, data extraction, and quality assessment processes, a second independent coder was not sought due to the relatively large scale of this work. Third, the present review protocol was not registered, and the search did not involve librarians. Despite these limitations, this article helps to resolve some of the controversies present in the literature while providing statistical support for the claims made in a previous narrative review (Harms et al., 2010). It also enhances our understanding of the specificity of facial emotion recognition deficits and the mechanisms underlying facial emotion recognition in ASD. Overall, this work provides some support for using measures of facial emotion recognition as a marker for ASD and informs intervention approaches to improve the emotion recognition skills—and hence the social competence—of autistic individuals.

Declaration of Competing Interest

The author declares no conflict of interest.

Acknowledgment

I would like to thank V. W. Chu and K. K. Chung for checking the codes.

Appendix A. Complete Keyword Strings

PubMed

("autis*"[All Fields] OR "asd"[All Fields]) OR ("asperger"[All Fields] OR "asperger s"[All Fields] OR "aspergers"[All Fields]) OR "PDD"[All Fields] OR ("child development disorders, pervasive"[MeSH Terms] OR ("child"[All Fields] AND "development"[All Fields] AND "disorders"[All Fields] AND "pervasive"[All Fields]) OR "pervasive child development disorders"[All Fields] OR ("pervasive"[All Fields] AND "developmental"[All Fields] AND "disorder"[All Fields]) OR "pervasive developmental disorder"[All Fields])) AND ("face"[MeSH Terms] OR "face"[All Fields] OR ("face"[MeSH Terms] OR "face"[All Fields] OR "facial"[All Fields] OR "facials"[All Fields])) AND ("emoting"[All Fields] OR "emotion s"[All Fields] OR "emotions"[MeSH Terms] OR "emotions"[All Fields] OR "emotion"[All Fields] OR "emotional"[All Fields] OR "emotive"[All Fields] OR ("affect"[MeSH Terms] OR "affect"[All Fields] OR "affects"[All Fields] OR "affected"[All Fields] OR "affecteds"[All Fields] OR "affecting"[All Fields]) OR ("express"[All Fields] OR "expresse"[All Fields] OR "expresses"[All Fields] OR "expressing"[All Fields] OR "expressions"[All Fields] OR "expressed"[All Fields] OR "expression"[All Fields]) OR "happ*"[All Fields] OR "sad"[All Fields] OR "ang"[All Fields] OR "disgust*"[All Fields] OR "surprise*"[All Fields] OR "fear*"[All Fields])

PsynINFO

((autis* or ASD or asperger or PDD or pervasive developmental disorder) and (face or facial) and (emotion or affect or expression or happ* or sad* or ang* or disgust* or surprise* or fear*)).mp. [mp = title, abstract, heading word, table of contents, key concepts, original title, tests & measures, mesh]

Appendix B. Quality Assessment Criteria based on the JBI Critical Appraisal Checklist for Case Control Studies

	Response			
	1 (Yes)	? (Unclear)	0 (No)	/ (N/A)
1. Were the groups comparable other than the presence of disease in cases or the absence of disease in controls?	Age, sex, and IQ were statistically comparable between ASD and TD groups (accepted if mental age and chronological age were matched using separate TD groups)	At least one of the variables was unreported AND other variables were comparable between groups	At least one of the variables was significantly different between groups	/
2. Were cases and controls matched appropriately?	Source population was the same for ASD and TD groups (e.g., autistic and TD individuals recruited from schools in the same city or state)	Source population was not mentioned or was ambiguous (e.g., participants recruited from the community) for at least one group	Source population was different for the two groups (e.g., autistic and TD individuals recruited from two different states or from hospitals and schools, respectively)	/
3. Were the same criteria used for identification of cases and controls?	Except for the ASD status, same inclusion/exclusion criteria (age, sex, or IQ) were stated for ASD and TD groups	At least one of the criteria was stated for one group only OR no criteria were stated for the two groups	Different criteria were stated for the two groups	/
4. Was exposure measured in a standard, valid and reliable way?	Individuals in ASD group met the diagnostic criteria using an ASD instrument (e.g., Autism diagnostic Interview-Revised) OR a diagnostic evaluation by at least two raters	Individuals in ASD group met the diagnostic criteria for ASD, but it is unclear whether an ASD instrument was involved OR a diagnostic evaluation was conducted by at least two raters	Not clear whether individuals in the ASD group met the diagnostic criteria for ASD AND neither an ASD instrument nor a diagnostic evaluation by at least two raters was involved	/
5. Was exposure measured in the same way for cases and controls?	A diagnostic evaluation, as well as an ASD/psychiatric instrument (if given to ASD group), were administered to TD individuals	Not clear whether a diagnostic evaluation, as well as an ASD/psychiatric instrument (if given to ASD group), were administered to TD group	Clear that no diagnostic evaluation or ASD/psychiatric instrument (if given to ASD group) was administered to TD group	/
6. Were confounding factors identified?	Age, sex, or IQ was matched or compared between ASD and TD groups	/	None of the variables was matched or compared between groups	/
7. Were strategies to deal with confounding factors stated?	ASD and TD groups were matched on age, sex, or IQ OR at least one of the variables was covaried or proven to have no significant relationship with the outcome variable	At least one of the variables was not reported for both groups AND none of the reported variables significantly differed between groups	At least one of the variables significantly differed between groups AND neither matching of groups nor covarying of variables was done	None of the variables significantly differed between groups AND neither matching of groups nor covarying of variables was done
8. Were outcomes assessed in a standard, valid and reliable way for cases and controls?	Published or validated FER tests and test stimuli were used AND the effect of ASD on the primary outcome (overall or individual emotion recognition performance) was assessed by $\geq$ 5 test trials (i.e., minimum number of trials needed to distinguish between above and below chance levels based on the binomial test)	Not clear whether the FER tests and test stimuli had been published or validated AND the effect of ASD on the primary outcome was assessed by $\geq$ 5 test trials	The effect of ASD on the primary outcome was assessed by $<$ 5 test trials	/
9. Was the exposure period of interest long enough to be meaningful?	/	/	/	/
10. Was appropriate statistical analysis used?	Statistical tests were checked for violation of test assumptions (e.g., normality and homogeneity of variance for *t*-tests and ANOVA) before analyzing the outcome variable OR nonparametric tests were used with justification	Statistical tests were not checked for violation of test assumptions before analyzing the outcome variable OR nonparametric tests were used without justification	Statistical tests were conducted even though there was a clear violation of test assumptions (e.g., $N < 10$ per parameter for linear regression OR the Events Per Variable value was $<$ 10 for logistic regression)	/

Note. ANOVA = analysis of variance; ASD = autism spectrum disorder; FER = facial emotion recognition; IQ = intelligence quotient; TD = typically developing. Item 9 was inapplicable because ASD is a lifelong condition. If the sex ratio, age range, or IQ range was the same for ASD and TD groups, criteria used for identification of cases and controls were assumed to be the same.

Appendix C. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.neubiorev.2021.104518.

References

Adolphs, R., 2002. Neural systems for recognizing emotion. Curr. Opin. Neurobiol. 12 (2), 169–177.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Baron-Cohen, S., Wheelwright, S., Jolliffe, A.T., 1997. Is there a" language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. Vis. Cogn. 4 (3), 311–331.

Barrett, L.F., Lindquist, K.A., Gendron, M., 2007. Language as context for the perception of emotion. Trends Cogn. Sci. (Regul. Ed.) 11 (8), 327–332.

Black, M.H., Chen, N.T., Iyer, K.K., Lipp, O.V., Bölte, S., Falkmer, M., et al., 2017. Mechanisms of facial emotion recognition in autism spectrum disorders: insights

from eye tracking and electroencephalography. Neurosci. Biobehav. Rev. 80, 488–515.
Bombari, D., Schmid, P.C., Schmid Mast, M., Birri, S., Mast, F.W., Lobmaier, J.S., 2013. Emotion recognition: the role of featural and configural face information. Q. J. Exp. Psychol. 66 (12), 2426–2442.
Bora, E., Pantelis, C., 2016. Meta-analysis of social cognition in attention-deficit/hyperactivity disorder (ADHD): comparison with healthy controls and autistic spectrum disorder. Psychol. Med. 46 (4), 699–716.
Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2011. Introduction to Meta-analysis. John Wiley & Sons.
Brosch, T., Pourtois, G., Sander, D., 2010. The perception and categorisation of emotional stimuli: a review. Cogn. Emot. 24 (3), 377–400.
Calvo, M.G., Lundqvist, D., 2008. Facial expressions of emotion (KDEF): identification under different display-duration conditions. Behav. Res. Methods 40 (1), 109–115.
Clark, T.F., Winkielman, P., McIntosh, D.N., 2008. Autism and the extraction of emotion from briefly presented facial expressions: stumbling at the first step of empathy. Emotion 8 (6), 803–809.
Cochran, W.G., 1954. The combination of estimates from different experiments. Biometrics 10 (1), 101–129.
Collin, L., Bindra, J., Raju, M., Gillberg, C., Minnis, H., 2013. Facial emotion recognition in child psychiatry: a systematic review. Res. Dev. Disabil. 34 (5), 1505–1520.
Darwin, C., 1872. The Expression of the Emotions in Man and Animals. John Murray., London.
Derntl, B., Seidel, E.M., Kainz, E., Carbon, C.C., 2009. Recognition of emotional expressions is affected by inversion and presentation time. Perception 38 (12), 1849–1862.
Doi, H., Fujisawa, T.X., Kanai, C., Ohta, H., Yokoi, H., Iwanami, A., et al., 2013. Recognition of facial expressions and prosodic cues with graded emotional intensities in adults with Asperger syndrome. J. Autism Dev. Disord. 43 (9), 2099–2113.
Drevon, D., Fursa, S.R., Malcolm, A.L., 2017. Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. Behav. Modif. 41 (2), 323–339.
Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56 (2), 455–463.
Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. Bmj 315 (7109), 629–634.
Enticott, P.G., Kennedy, H.A., Johnston, P.J., Rinehart, N.J., Tonge, B.J., Taffe, J.R., Fitzgerald, P.B., 2014. Emotion recognition of static and dynamic faces in autism spectrum disorder. Cogn. Emot. 28 (6), 1110–1118.
Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Bölte, S., et al., 2016. Basic and complex emotion recognition in children with autism: cross-cultural findings. Mol. Autism 7 (1), 52.
Gepner, B., Deruelle, C., Grynfeltt, S., 2001. Motion and emotion: a novel approach to the study of face processing by young autistic children. J. Autism Dev. Disord. 31 (1), 37–45.
Goldman, A.I., Sripada, C.S., 2005. Simulationist models of face-based emotion recognition. Cognition 94 (3), 193–213.
Gross, T.F., 2004. The perception of four basic emotions in human and nonhuman faces by children with autism and other developmental disabilities. J. Abnorm. Child Psychol. 32 (5), 469–480.
Happé, F., Frith, U., 2006. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. J. Autism Dev. Disord. 36 (1), 5–25.
Harms, M.B., Martin, A., Wallace, G.L., 2010. Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychol. Rev. 20 (3), 290–322.
Hartung, J., Knapp, G., 2001. A refined method for the meta-analysis of controlled clinical trials with binary outcome. Stat. Med. 20 (24), 3875–3889.
Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21 (11), 1539–1558.
Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., Traue, H.C., 2010. Expression intensity, gender and facial emotion recognition: women recognize only subtle facial emotions better than men. Acta Psychol. (Amst) 135 (3), 278–283.
Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., Youngstrom, E., 2001. Emotion knowledge as a predictor of social behavior and academic competence in children at risk. Psychol. Sci. 12 (1), 18–23.
Kätsyri, J., Saalasti, S., Tiippana, K., von Wendt, L., Sams, M., 2008. Impaired recognition of facial emotions from low-spatial frequencies in Asperger syndrome. Neuropsychologia 46 (7), 1888–1897.
Lau, J.Y., Burt, M., Leibenluft, E., Pine, D.S., Rijsdijk, F., Shiffrin, N., Eley, T.C., 2009. Individual differences in children's facial expression recognition ability: the role of nature and nurture. Dev. Neuropsychol. 34 (1), 37–51.
Law Smith, M.J.L., Montagne, B., Perrett, D.I., Gill, M., Gallagher, L., 2010. Detecting subtle facial emotion recognition deficits in high-functioning autism using dynamic stimuli of varying intensities. Neuropsychologia 48 (9), 2777–2781.
Lindquist, K.A., Barrett, L.F., Bliss-Moreau, E., Russell, J.A., 2006. Language and the perception of emotion. Emotion 6 (1), 125–138.
Lozier, L.M., Vanmeter, J.W., Marsh, A.A., 2014. Impairments in facial affect recognition associated with autism spectrum disorders: a meta-analysis. Dev. Psychopathol. 26 (4pt1), 933–945.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. 151 (4), 264–269.
Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., Mu, P.-F., 2020. Chapter 7: systematic reviews of etiology and risk. In: Aromataris, E., Munn, Z. (Eds.), JBI Manual for Evidence Synthesis. JBI. Available from. https://synthesismanual.jbi.global.
Nook, E.C., Lindquist, K.A., Zaki, J., 2015. A new look at emotion perception: concepts speed and shape facial emotion recognition. Emotion 15 (5), 569.
Piggot, J., Kwon, H., Mobbs, D., Blasey, C., Lotspeich, L., Menon, V., et al., 2004. Emotional attribution in high-functioning individuals with autistic spectrum disorder: a functional imaging study. J. Am. Acad. Child Adolesc. Psychiatry 43 (4), 473–480.
Rhodes, G., Burton, N., Jeffery, L., Read, A., Taylor, L., Ewing, L., 2018. Facial expression coding in children and adolescents with autism: reduced adaptability but intact norm-based coding. Br. J. Psychol. 109 (2), 204–218.
Rohatgi, A., 2019. WebPlotDigitizer. Retrieved from. https://automeris.io/WebPlotDigitizer/.
Schirmer, A., Adolphs, R., 2017. Emotion perception from face, voice, and touch: comparisons and convergence. Trends Cogn. Sci. 21 (3), 216–228.
Schwarzer, G., Carpenter, J.R., Rücker, G., 2015. Meta-Analysis With R, Vol. 4784. Springer, Cham.
Sidik, K., Jonkman, J.N., 2002. A simple confidence interval for meta-analysis. Stat. Med. 21 (21), 3153–3159.
Thalheimer, W., Cook, S., 2002. How to calculate effect sizes from published research: a simplified methodology. Work-Learning Res. 1, 1–9.
Uljarevic, M., Hamilton, A., 2013. Recognition of emotions in autism: a formal meta-analysis. J. Autism Dev. Disord. 43 (7), 1517–1526.
Weigelt, S., Koldewyn, K., Kanwisher, N., 2012. Face identity recognition in autism spectrum disorders: a review of behavioral studies. Neurosci. Biobehav. Rev. 36 (3), 1060–1084.
Yeung, M.K., Lee, T.L., Chan, A.S., 2019. Impaired recognition of negative facial expressions is partly related to facial perception deficits in adolescents with high-functioning autism spectrum disorder. J. Autism Dev. Disord. 1–11.