Review

autism

Autism
2018, Vol. 22(2) 99–117

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1362361316668652
journals.sagepub.com/home/aut
SAGE

Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review

Tamar Kalandadze[1], Courtenay Norbury[2], Terje Nærland[3] and Kari-Anne B Næss[1]

Abstract

We present a meta-analysis of studies that compare figurative language comprehension in individuals with autism spectrum disorder and in typically developing controls who were matched based on chronological age or/and language ability. A total of 41 studies and 45 independent effect sizes were included based on predetermined inclusion criteria. Group matching strategy, age, types of figurative language, and cross-linguistic differences were examined as predictors that might explain heterogeneity in effect sizes. Overall, individuals with autism spectrum disorder showed poorer comprehension of figurative language than their typically developing peers (Hedges' $g = -0.57$). A meta-regression analysis showed that group matching strategy and types of figurative language were significantly related to differences in effect sizes, whereas chronological age and cross-linguistic differences were not. Differences between the autism spectrum disorder and typically developing groups were small and nonsignificant when the groups were matched based on the language ability. Metaphors were more difficult to comprehend for individuals with autism spectrum disorder compared with typically developing controls than were irony and sarcasm. Our findings highlight the critical role of core language skills in figurative language comprehension. Interventions and educational programmes designed to improve social communication skills in individuals with autism spectrum disorder may beneficially target core language skills in addition to social skills.

Keywords

autism spectrum disorder, figurative language, meta-analysis, metaphor

Introduction

Autism spectrum disorder (ASD) is a pervasive developmental disorder that is characterized by impairments in social interaction and social communication, and by restricted, repetitive behaviours and interests (American Psychiatric Association (APA), 2013). While the relationship between ASD and core language skills has been debated for several years, many studies have consistently reported that the social use of language, particularly figurative language (the ability to go beyond what is explicitly stated), is universally impaired in individuals with ASD (e.g. Dennis et al., 2001; Happé, 1993, 1994). In particular, literal interpretations of utterances with intended nonliteral meanings have been considered characteristic of these individuals (e.g. Happé, 1993; MacKay and Shaw, 2004).

Despite this level of consensus, a growing body of research indicates that a subset of individuals with ASD do not differ significantly from typically developing (TD) controls with similar language ability (LA) in terms of selected measures of figurative language (Hermann et al., 2013; Norbury, 2004, 2005). Even in studies in which participants with ASD receive lower scores than their TD controls, performance is often above chance level (e.g. Wang et al., 2006), indicating that individuals with ASD are not consistently inclined towards literal meanings. These

[1]University of Oslo, Norway
[2]University College London, UK
[3]Oslo University Hospital, Norway

Corresponding author:
Tamar Kalandadze, Department of Special Needs Education, University of Oslo, PO Box 1140, Blindern, 0318 Oslo, Norway.
Email: tamar.kalandadze@isp.uio.no

findings indicate that deficits in figurative language comprehension may not be unique or/and universal among individuals with ASD, fuelling debates regarding the source and the extent of the difficulty associated with figurative language comprehension in people with ASD.

In this article, we present a meta-analysis of studies that have explicitly compared individuals with ASD with TD controls in terms of their figurative language comprehension. We focus on whether (a) the group matching strategy (i.e. whether the ASD and TD groups are matched based on chronological age (CA), LA or both), (b) differences with regard to CA, (c) the type of figurative language (*tropes*) measured and (d) cross-linguistic differences (the languages in which the studies are conducted) can explain between-study variance. A better understanding of these factors may have far-reaching implications for education and clinical interventions involving individuals with ASD.

What is figurative language?

Generally, in figurative language, the intended meanings of the words, sentences and expressions used do not coincide with their literal meanings (Gluksberg, 2001). When speaking figuratively, speakers mean something other than what they literally say (Gibbs and Colston, 2012). Therefore, to understand figurative language, an individual must be able to grasp the speaker's intention in a given context (Rapp and Wild, 2011).

The most common examples of figurative language include *metaphors* (e.g. 'Love is a journey'), which involves 'understanding and experiencing one kind of thing in terms of another' (Lakoff and Johnson, 2003: 5), and *verbal irony* (e.g. 'What nice weather' – to describe a rainy day), which 'consists in echoing a thought (e.g., a belief, an intention, a norm-based expectation) attributed to an individual, a group or to people in general, and expressing a mocking, skeptical or critical attitude to this thought' (Wilson and Sperber, 2012: 125). A sub-type of irony is *sarcasm*, which is often intended to insult or wound (Rapp and Wild, 2011) (e.g. 'I just love your blouse', said by someone who does not actually like the blouse).

Traditionally, figurative language has been regarded as deviant of so-called literal language. By contrast, it is now well established that figurative language is a ubiquitous part of daily language and social communication. For instance, approximately 8% of adult utterances in conversations among friends contain some irony (Gibbs, 2000), teachers frequently use figurative language when presenting their lessons to students (Kerbel and Grunwell, 1997), and literature is replete with metaphors (Colston and Kuiper, 2002). Consequently, figurative language comprehension influences social relationships, social participation and educational achievement (Cain et al., 2005; Kerbel and Grunwell, 1997; Nippold, 1991; Swineford et al., 2014). As a result, deficits in figurative language comprehension may seriously affect an individual's life.

Figurative language comprehension in TD individuals

In TD individuals, the ability to comprehend figurative language emerges in early childhood, continues to develop steadily into adolescence, and improves throughout adulthood (e.g. Ackerman, 1982; Dews et al., 1996; Falkum et al., 2016; Hancock et al., 2000; Pexman et al., 2005; Pexman and Glenwright, 2007; Rundblad and Annaz, 2010a; Semrud-Clikeman and Glass, 2010; Winner, 1988; Winner et al., 1976). However, findings vary regarding the age at which the ability to comprehend different tropes emerges, which may partially be due to theoretical and methodological inconsistencies across studies (Pouscoulous, 2011; Winner et al., 1976).

The developmental achievements that underpin figurative language comprehension are the subject of theoretical debate. Some claim that figurative language develops in parallel with the acquisition of a theory of mind (ToM: the capacity to attribute one's own mental states and those of others) due to the presumed need to appreciate a speaker's intended message, which may not be literal (Happé, 1993). Pouscoulous (2011) suggests that the cognitive means for understanding figurative language (e.g. metaphor) are present by the time a child starts to speak. As children age, their language skills, world knowledge, and cultural expertise develop, which may account for their improved figurative language comprehension. Importantly, because figurative language tropes differ in terms of structure and processing demands (Colston and Gibbs, 2002), they seemingly require different mechanisms to be understood.

Notably, when considering the role of ToM and/or language skills in figurative language comprehension, the well-documented close relationships between TD children's linguistic and communicative skills (semantic, syntax and pragmatics) and ToM make teasing out the independent contributions of ToM and language skills in figurative language comprehension difficult. For instance, accuracy in standard false-belief tasks has been found to be strongly correlated with participants' receptive and expressive language abilities (Pons et al., 2009).

Figurative language comprehension in individuals with ASD

A number of studies have shown that individuals with ASD have difficulty understanding figurative language (e.g. Happé, 1993; Kaland et al., 2002; Rundblad and Annaz, 2010b). Two primary explanations for such difficulties have been proposed. First, the social cognitive profile that is suggested to be characteristic of individuals with ASD, including deficits in ToM development (Baron-Cohen et al., 1985), is thought to account for specific deficits in their figurative language comprehension. Happé (1993) conducted the first experimental research to make explicit connections between ToM and figurative language comprehension in individuals with ASD. She examined relevance theory (Sperber and Wilson, 1986) and ToM explanations for individuals with ASD and found that only the individuals with ASD who passed the first-order ToM tasks (inferring a person's mental state, e.g. what he/she thinks) performed well on metaphor tasks, whereas performance on irony tasks required that these individuals be capable of passing second-order ToM tasks (considering embedded mental states (Baron-Cohen, 2001), e.g. what he thinks that she thinks). The study concluded that ToM understanding predicted performance on metaphor and irony tasks.

Second, figurative language comprehension deficits are neither universal nor specific to individuals with ASD (Gernsbacher and Pripas-Kapit, 2012), but they can be related to an individual's structural language skills (vocabulary and syntax) (Norbury, 2004, 2005; Whyte et al., 2014). For example, Norbury (2004) found that children and adolescents with ASD did not show impairments in figurative language comprehension compared with TD controls when the former's structural language abilities (vocabulary and syntax) were within the normal range.

This hypothesis is supported by recent findings that reveal impairments in structural language skills (e.g. syntax) in some linguistically able individuals with ASD (Brynskov et al., 2016; Eigsti et al., 2011). Given the evidence that syntax is one of the most important predictors of success in, for example, metaphor comprehension in TD individuals (Pouscoulous, 2014), the impairments in syntactic ability in many individuals with ASD may affect their figurative language comprehension (e.g. Whyte et al., 2014).

Additionally, comparisons between distinct neurocognitive phenotypes within ASD reveal different patterns of language comprehension. More specifically, the autism language impairment (ALI) phenotype includes individuals with ASD who meet standard diagnostic criteria for language impairment (Tager-Flusberg, 2006; Tager-Flusberg and Joseph, 2003). The problems that individuals with ALI experience in comprehending structural language vary, and they are not unique to ASD or necessarily related to the severity of core ASD symptoms or overall cognitive functioning (Tager-Flusberg and Joseph, 2003). When individuals with ALI are compared with individuals with ASD, who score within the normal range on standard language tests (Autism language normal (ALN)), differences in figurative language comprehension are evident (Gernsbacher and Pripas-Kapit, 2012; Norbury, 2004, 2005).

Gernsbacher and Pripas-Kapit (2012) argued that, when researchers control for language comprehension in ToM tasks, differences in figurative language comprehension between ASD and TD groups disappear. Similarly, studies have shown that core language skills remain a significant predictor of figurative language comprehension in individuals with ASD after ToM has been considered (Norbury, 2005).

In addition to these two primary explanations about the involvement of ToM or core/structural language skills in figurative language comprehension, several studies have concluded that both ToM and language (syntax) are uniquely related to figurative language comprehension (idioms) in individuals with ASD (e.g. Whyte et al., 2014). However, given that relatively few studies have examined ToM ability alongside language skills in individuals with ASD, the extent to which language skills and ToM independently contribute to figurative language comprehension remains to be explained.

Poor figurative language comprehension seems to be a persistent challenge for individuals with ASD (De Villiers et al., 2011; MacKay and Shaw, 2004), with deficits reported in adults as well (e.g. Ozonoff and Miller, 1996). However, due to the paucity of longitudinal studies investigating figurative language development in individuals with ASD, conclusions on the developmental path of this skill in ASD are difficult to draw. A cross-sectional study by Whyte and Nelson (2015) found that, for 7- to 12-year-old children with ASD, performance on nonliteral language tasks increased significantly with CA. Therefore, we hypothesize that CA may be an important determinant of figurative language comprehension in individuals with ASD, as it is in TD individuals.

Figurative language comprehension deficits in individuals with ASD have been reported in studies conducted in different languages (e.g. Adachi et al., 2004; Kaland et al., 2002). Given the differences between languages and the influence of culture on language, figurative language may differ depending on the language used. However, until now, most studies have been conducted in English-speaking countries, which may lead to language and cultural biases.

In summary, there is a significant gap in our knowledge regarding the extent of figurative language comprehension deficits in individuals with ASD relative to TD controls. Moreover, there is little consensus regarding the factors that underlie figurative language abilities in individuals with ASD – a situation that is exacerbated by inconsistent and contradictory findings in the literature. Resolving these issues is crucial to identify appropriate targets and methods for interventions.

Previous reviews

Although several narrative reviews have focused on figurative language comprehension in individuals with ASD (e.g. Gernsbacher and Pripas-Kapit, 2012; Hobson, 2012; Lyons and Fitzgerald, 2004; Melogno et al., 2012; Passanisi and Di Nuovo, 2015; Pexman, 2008; Samson, 2013; Vulchanova et al., 2015), no prior meta-analysis has summarized group differences and similarities in figurative language comprehension between individuals with ASD and TD individuals.

The current study

The present meta-analytic review examines (a) the differences and similarities between individuals with ASD and TD controls with regard to figurative language comprehension and (b) variables (group matching strategy, CA, tropes, and cross-linguistic differences) that may explain the differences in the results across studies.

Given that research on figurative language comprehension in individuals with ASD often involves small samples, which can threaten statistical power (Næss et al., 2011), meta-analyses are useful because they statistically aggregate study findings, provide the effect sizes of group differences and thus offer increased statistical power (Borenstein et al., 2009).

Method

To ensure methodological quality, the present meta-analysis was designed and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (www.prisma-statement.org).

Literature search

The electronic database search was conducted under a University Librarian's supervision. First, we searched through electronic databases (PsychInfo, Eric, Embase, Web of Science, Medline, ScienceDirect, Linguistics and Language Behavior Abstracts (LLBA) and NoRART) for studies published until February 2016, using combinations of keywords related to ASD – *ASD, Asperger Syndrome or asperger**, *Autism or autis**, and *Pervasive Developmental Disorders or* '*pervasive developmental disorder**' – crossed with keywords related to figurative language – *Figurative Language or* '*figurative language*', '*figure** *of speech*', *humor or humour, hyperbole**, *idiom**, '*indirect speech*', *irony, metaphor**, *metonymy*, '*non-literal language*', *sarcasm* and *simile*. Second, we examined the references of the included studies as well as relevant narrative reviews to retrieve additional sources that were not already included in the search results. Third, we manually searched through issues of the *Autism, Autism Research*, the *Journal of Autism and Developmental Disorders, Research in Developmental Disabilities*, and the *International Journal of Language and Communication Disorders* as well as the Publications in Research page on autism. We also examined a special issue of *Metaphor and Symbol* (2012; 27(1)) that was devoted to ASD. Fourth, to minimize potential publication bias, we searched for grey literature and also emailed key authors in the field, asking them for unpublished studies. Including grey literature is important as, in general, evidence shows that studies that report large effect sizes and significant results are more easily published than studies that report null findings or small effect sizes (Borenstein et al., 2009).

Study inclusion criteria

Selected articles were required to meet the following predetermined inclusion criteria: (1) the results of any trope(s) of figurative language comprehension, measured with either standardized or nonstandardized assessment tasks, had to be reported. Several studies involved one or more measures of figurative language tropes as an aspect of ToM, but only pure measures of figurative language were included in this study. When insufficient data were available to calculate the effect sizes and when necessary data could not be obtained from the author(s), studies were excluded. (2) Participants had to be diagnosed with ASD using the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) or *International Classification of Diseases* (ICD) diagnostic criteria, and they had to have a verbal IQ (VIQ) score of >70. Thus, studies that examined participants with normal intelligence (i.e. high-functioning ASD (HFASD), high-functioning autism (HFA), or Asperger syndrome or Pervasive Developmental Disorder–Not Otherwise Specified (PDD-NOS)) were included. (3) A comparison group of CA- and/or LA-matched TD controls had to be included, and (4) studies had to be reported in English, Russian, Norwegian, Swedish or Danish because at least one of the authors master one or more of these languages.

Screening process

The first author and a research assistant judged the relevance of the abstracts obtained from the search results. When an abstract contained insufficient information, the full-text article was reviewed. The papers that met the inclusion criteria based on the abstracts were examined. Finally, 41 studies were coded. For further information on the screening process and the reasons that studies were excluded, see Figure 1.

Coding

Study characteristics. Study characteristics (title, author(s) and publication year) were coded for descriptive purposes. For the main analysis, we coded the number of ASD participants and the number of TD participants in addition to inferential statistics based on means and standard deviations (SDs), mean rates of correct answers, *p*-values or chi-square values. Because computing an overall effect

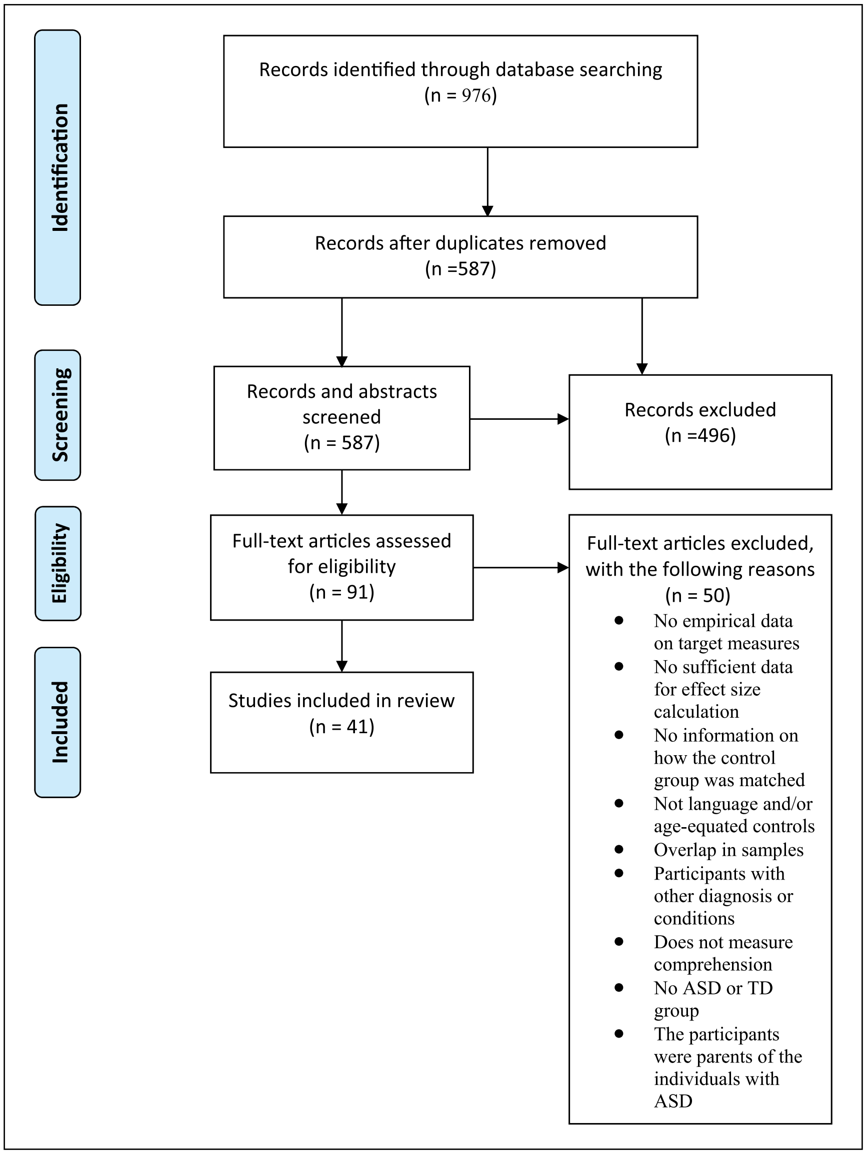

Figure 1. Flow diagram for the search and inclusion of studies.

size more than once based on information from the same sample can lead to incorrect estimates (Borenstein et al., 2009), for the studies that included multiple data collection points (e.g. intervention studies), only the first data collection point was coded. The first time point usually provides the largest sample due to attrition over time and because the results may not be influenced by any intervention effects; in descriptive studies with overlapping samples, the data from the study with the largest sample were coded; for studies that included measures of neuroimaging, only behavioural results were coded.

Predictor variables. The following predictor variables were selected and coded.

Age. The mean ages of the participants with ASD and the participants with TD were coded.

Group matching strategy. We coded whether the ASD and TD groups were matched based on CA, LA or both. Details about group matching variables are presented in Appendix 1.

Tropes. All figurative language tropes that were examined in the included studies were coded.

Cross-linguistic differences. The languages in which the studies were conducted were coded into two categories: (1) the English language (26 studies) and (2) other languages: Hebrew (4 studies), Japanese (2 studies), Taiwanese (2 studies), German (1 study), Danish (1 study), Korean (1 study), French (1 study), Chinese (1 study), Cantonese (1 study) and Dutch (1 study).

We also planned to use ToM as a predictor variable, but, due to the small number of studies that examined ToM independently of figurative language ($n=7$; Adachi et al., 2004; De Villiers et al., 2011; Huang et al., 2015; Martin and McDonald, 2004; Norbury, 2004; Rundblad and Annaz, 2010b; Whyte et al., 2014), no analysis including ToM could be conducted.

Multiple subgroups and multiple outcomes within studies. The included studies involve complex data structures: (a) *independent subgroups* within studies (e.g. children, adolescents and adults), (b) *dependent subgroups* within studies (e.g. one target group compared with two TD control groups or two target groups compared with one TD control group) and (c) *multiple outcomes* within studies (e.g. various tropes of figurative language or various forms of the same trope, such as novel and conventional metaphors).

Treating dependent effect sizes as independent introduces bias by giving more weight in the meta-analysis to the studies with multiple outcomes or more than two groups (Scammacca et al., 2014). To avoid this problem, several considerations were made prior coding. All the *independent and dependent subgroups* within the included studies were coded and they could contribute either as separate scores or as composite scores in the analysis. The choices made with regard to treating the information that comes from independent and dependent groups are explained in detail in the section on the meta-analytic procedure and analysis.

Two independent raters coded 100% of the included studies: the first author of this article and a doctoral candidate who is trained in meta-analysis. The inter-rater Cronbach's alpha was $\alpha=0.99$ for CA, $\alpha=1.00$ for tropes, and $\alpha=0.98$ for the outcome measures. The disagreements between the raters were resolved by consulting original papers and/or via discussions between the raters.

Meta-analytic procedures and analysis

All statistical analyses were conducted using *Comprehensive Meta-Analysis* (CMA) Software, version 3 (Biostat). The data reported in different formats were combined in the same analysis because CMA computes effect size from different formats.

The effect sizes for all studies were computed using Hedges' g, which is corrected for small sample sizes and thus tends to be unbiased for small samples (Hedges, 1981). Cohen's general guidelines were used to judge the effect size. Based on the suggestions from Cohen (1968), effect sizes were referred as small ($d=0.2$), medium ($d=0.5$) and large ($d=0.8$). However, note that such tentative guidelines should be interpreted with caution and preferably in relation to the practical consequences that the effect size may have (Lakens, 2013). When Hedges' g value was positive, individuals with ASD had the highest group mean; when its value was negative, the group difference favoured the TD group. A 95% confidence interval (CI) was calculated for each effect size to indicate whether it was statistically significantly greater than zero. The effect is statistically significant if the CI does not cross zero.

The overall effect size was estimated by calculating a weighted average of individual effect sizes. Effect size calculations were based on a random-effects model, which, unlike the fixed-effects model, assumes that between-study variations in effect size not only result from random error but also are systematic and that the variations depend on variables, which are likely to vary from study to study (Borenstein et al., 2009). The assumption that effect sizes are heterogeneous suggests that other factors beyond an ASD diagnosis have an impact on the differences in effect sizes between studies. In particular, the effect sizes were predicted to vary based on CA, group matching strategy, tropes, and the languages in which the included studies were conducted.

Heterogeneity

We examined between-study heterogeneity in effect-size distributions and degrees of homogeneity using Q and I^2 (Hedges and Olkin, 1985). The Q-statistic index shows heterogeneity in effect size across the studies, and I^2 explains that a proportion of the variance is real and not simply due to random error (Borenstein et al., 2009). A significant result on this test indicates significant heterogeneity between the studies' effect sizes.

Publication bias

A funnel plot was used to determine whether publication bias was present. A symmetrical funnel would indicate the absence of publication bias.

Multiple subgroups and multiple outcomes analysis

The *independent subgroups* within the studies were treated as separate studies because they provide independent/unique information from the different participants. By contrast, the *dependent subgroups* do not provide unique information because the same participants are included in each comparison and the resulting effect sizes are statistically dependent. The inclusion of statistically dependent effect sizes in a meta-analysis can present a serious threat to its validity (Borenstein et al., 2009; Scammacca et al., 2014). First, the inclusion of dependent effect sizes leads to incorrect estimates of the variance in the summary effect; second, in the meta-analysis, it gives more weight to studies that have multiple measures. Therefore, in this study, the results from the *dependent subgroups* were aggregated as a composite score, contributing to only one effect size and thus eliminating the dependence from, for example, the shared control group.

Creating a composite score in our meta-analysis resulted in combining the measures from the dependent subgroups that compared either one ASD to two TD groups – one matched based on LA and another matched based on CA – or compare a TD group with both ASL (ASD with language impairment) and ASO (ASD only – ASD without language impairment) groups (Norbury, 2004). Although it is considered an appropriate way to treat dependent measures, this approach obscures important findings in these studies that are directly relevant to this meta-analysis. Namely, when the ASD and TD groups are matched based on LA, inter-group differences are usually very small and nonsignificant in contrast to the findings of a comparison of the same ASD group with a CA-matched TD group. Similarly, the group difference is very small and nonsignificant when the TD group is compared with the ASO group, and large and significant when the TD group is compared with the ASL group (Norbury, 2004).

Similar to the case of dependent subgroups, all *multiple outcomes* were collapsed into a composite score, thereby contributing to only one effect size. The composite score is defined as the mean effect size in a study, with a variance that considers the correlation among the different outcomes. Thus, every study is represented by one score in the main analysis, regardless of the number of outcomes or dependent subgroups included in the mean.

Meta-regression analysis

Random-effects models were used in the meta-regression analysis to avoid the problem of the difference in the weighting of small studies and large studies in the fixed-effects models. Using random-effects models ensures that each study is weighted more evenly, irrespective of the study's sample size. To determine the strength of the predictors of the outcomes, R^2 (a percentage of the explained between-study variance) was used as an effect size. R^2 index quantifies the proportion of variance explained by the covariates (Borenstein et al., 2009).

Analysis by subgroups

To examine how the differences in effect sizes across studies varied according to the group matching strategies, we ran an *analysis by subgroup* that involved the following matching variables: CA, LA, or both CA and LA. Two studies (Pexman et al., 2011; Whyte et al., 2014) were excluded from this analysis because they included a dependent ASD group and two TD groups, and, as explained earlier, a composite score of the results was created.

Analysis by outcomes – metaphors and irony/sarcasm

Given the evidence that different tropes of figurative language are likely to be comprehended differently by individuals with ASD (Happé, 1993), we compared the studies involving metaphor with the studies involving irony and sarcasm (combined). These tropes were chosen because they were the most frequently studied tropes in the included studies.

To examine group differences across these tropes, we used the *select by outcome analysis* function of CMA. Three studies (Adachi et al., 2004; De Villiers et al., 2011; Huang et al., 2015) examined both metaphor and irony or sarcasm and were therefore included only with metaphor studies because their inclusion in both groups would lead to validation problem due to statistically dependent effect sizes. They were included with the metaphor studies because there were fewer metaphor studies compared with irony and sarcasm studies.

Two studies did not differentiate metaphors from idioms and figures of speech (Dennis et al., 2001; Landa and Goldberg, 2005); therefore, they were excluded from the analysis. Overall, 13 independent effect sizes examined different types of metaphor, and 20 independent effect sizes examined irony and sarcasm.

Results

The characteristics of each study included in the meta-analysis are presented in Appendix 1. Figure 2 shows the group differences with CIs between individuals with ASD and TD individuals in terms of the comprehension of figurative language.

Figurative language comprehension in individuals with ASD compared with that in TD controls

A total of 45 independent effect sizes, involving 1119 individuals with ASD (mean sample size = 27.97, SD = 21.12, range = 8–164) and 978 TD controls (mean sample size = 24.45, SD = 15.77, range = 8–164), examined the differences in figurative language comprehension between the two groups. The standardized mean effect size was moderate and significant (g = –0.57, 95% CI = [–0.72, –0.41], $p < 0.001$) in favour of TD individuals. The heterogeneity between studies was significant (Q (44) = 138.56, $p = < 0.001$, $I^2 = 68.24$). The funnel plot examining the publication bias in the analysis showed symmetrical distribution, indicating no publication bias.

Impact of age, tropes, matching strategy and cross-linguistic differences

The meta-regression analysis including group matching strategy, age, different tropes, and cross-linguistic differences generated a significant result (Q (9) = 28.85; $p < 0.001$), indicating that effect size is related to at least one of the covariates. The model reliably explained 41% of the variance in the effect sizes between the studies.

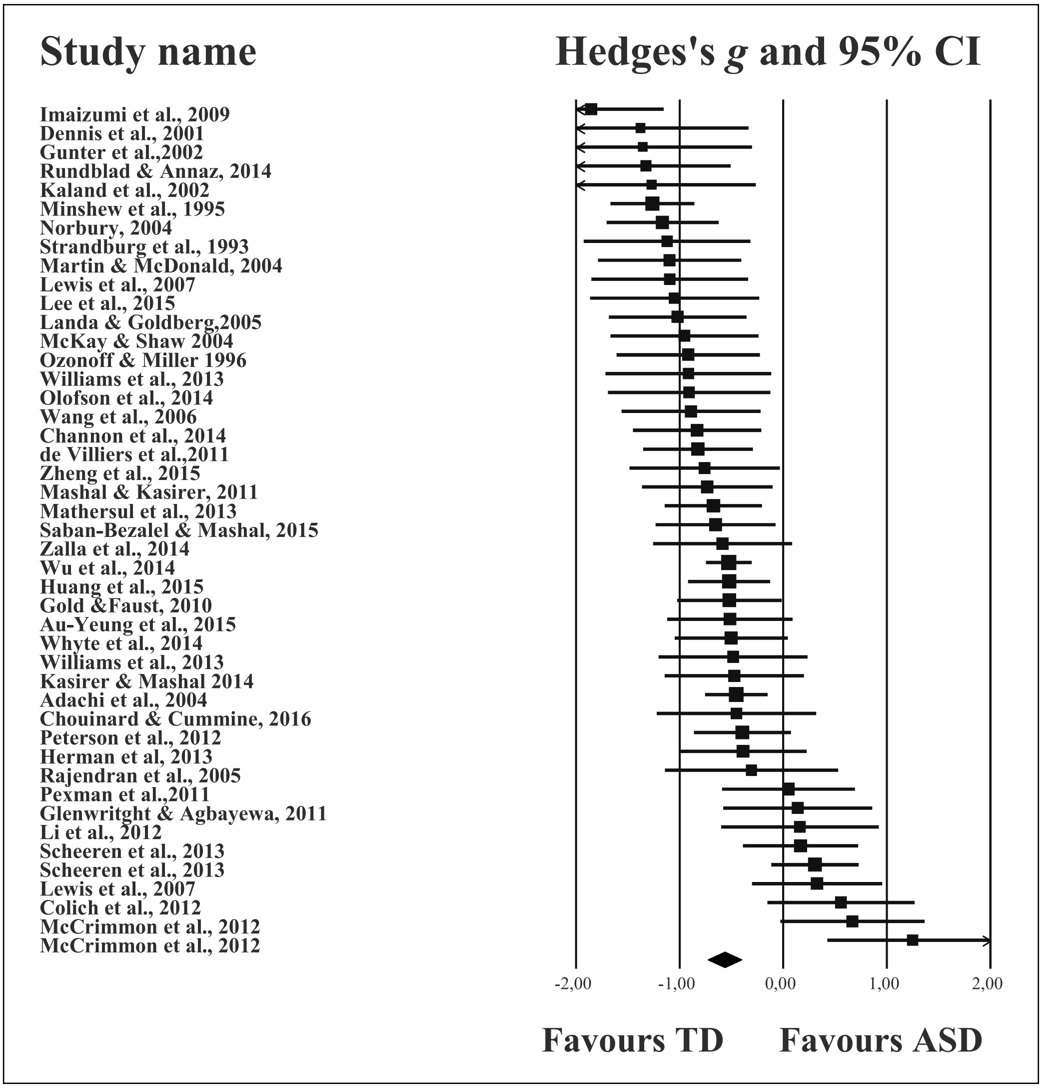

Figure 2. Overall mean effect size for group differences in figurative language comprehension comparing individuals with ASD and TD controls (Hedges' *g*, displayed by ♦) and effect sizes with confidence intervals for each study represented by horizontal lines.

However, an examination of the impact of each covariate revealed that age and cross-linguistic differences do not have a significant impact on the differences in effect sizes across studies ($p = 0.140$; $p = 0.543$, respectively). By contrast, tropes and group matching strategy were significantly related to the effect sizes ($p < 0.002$ and $p < 0.035$, respectively).

Although we could not include ToM as an independent measure due to an insufficient number of studies, a narrative summary of the results showed that some of the studies that measured ToM independently from figurative language tropes found correlations between ToM and one or more trope (Adachi et al., 2004; De Villiers et al., 2011; Martin and McDonald, 2004; Whyte et al., 2014), while others did not (Norbury, 2005; Rundblad and Annaz, 2010b). Huang et al. (2015) found that ToM understanding was partially related to figurative language comprehension in children with ASD. Namely, children with no ToM competence showed lower scores in figurative language comprehension than children who achieved first-order or second-order ToM. However, no differences were found between first-order ToM achievers and second-order ToM achievers. Notably, different ToM tasks have been used across these studies, which could have resulted in different findings in relation to figurative language comprehension.

Impact of group matching strategy

The results of the *analysis by subgroups* showed that mean effect sizes varied across the studies depending on whether the ASD and TD groups were matched based on CA, LA or both ($g = -0.92$, $g = -0.06$ and $g = -0.55$, respectively). In particular, the group differences in studies that included LA-matched groups were small and nonsignificant, they were large in studies that included CA-matched groups, and they were moderate in studies that included both CA- and LA-matched groups. The test to compare the difference in the results for these matching variables yielded a Q-value of 10.33 with $3df$ and $p=0.016$, indicating that the group matching strategy significantly influences between-study differences.

Narrative summary of the comparisons of the dependent subgroups. Due to the small number of the studies that compared more than two dependent subgroups ($n=3$; Norbury, 2004; Pexman et al., 2011; Whyte et al., 2014), we could not conduct a quantitative analysis. Instead, we report a brief narrative summary of the results of these studies.

Norbury (2004) found large group difference in the figurative language (idioms) comprehension of the TD group and the groups with children with ASL. However, the differences between children with ASO and the TD controls were small and nonsignificant. The children were assessed with the British Picture Vocabulary Scale (BPVS II; Dunn et al., 1997), the Concepts and Directions Subtest of the Clinical Evaluation of Language Fundamentals (CELF-III; Semel et al., 2000), and the Recalling Sentences Subtest of the CELF-III (Conti-Ramsden et al., 2001).

Pexman et al. (2011) found small differences between the CA-matched ASD and TD groups. Even smaller difference in favour of the participants with ASD was found between individuals with ASD and their LA-matched TD counterparts. The picture vocabulary subscale of the Test of Language Development–Primary, Third Edition (TOLD-P: 3; Newcomer and Hammill, 1997) was used as an LA measure to match the ASD and TD groups.

Whyte et al. (2014) showed that the ASD group had more deficits in figurative language (idioms) comprehension than did the CA-matched TD group. However, when the same ASD group was compared to another LA-matched TD group, the group difference was found to be small in favour of ASD. The participants in the LA-matched ASD and TD groups were matched based on their syntactic abilities as assessed with Syntax Construction subtest of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999).

Impact of tropes – metaphor and irony/sarcasm. The *select by outcome analysis* comparing TD and ASD groups on measures of metaphor, on the one hand, and irony and sarcasm, on the other, showed that group differences were larger in studies that measured metaphor comprehension (Hedges' $g=-0.72$) compared with the studies that measured irony and sarcasm (Hedges' $g=-0.48$).

Discussion

In this study, we investigated figurative language comprehension in individuals with ASD compared with CA- or/ and LA-matched TD controls. In addition, we examined variables that could reliably explain variations in effect size across studies. The results showed that (1) on average, individuals with ASD fall behind in their comprehension of various figurative language tropes relative to TD peers; (2) the between-study differences can reliably be explained by the group matching strategy and the trope measured; (3) age and cross-linguistic differences cannot account for between-study differences. These findings and their implications for future research and practice are discussed in the following sections.

Figurative language comprehension is challenging for many individuals with ASD

Overall, individuals with ASD exhibited moderately poorer figurative language comprehension skills compared with their TD controls. This finding is consistent with a number of studies (e.g. Happé, 1993; Kaland et al., 2002) and indicates that the social communication problems that individuals with ASD generally have may partially covary with their poor understanding of figurative language. However, the studies that included LA-matched groups displayed nonsignificant differences between the ASD and TD groups. This finding indicates that the figurative language deficit is seemingly neither universal nor unique to individuals with ASD; instead, it appears to be related to participants' language skills. This finding will be discussed in the next section.

Is figurative language comprehension related to core language skills in individuals with ASD?

Although individuals with ASD, in general, showed greater difficulties in understanding figurative language than TD controls, a high level of heterogeneity in the effect sizes across the studies examined indicates that factors other than an ASD diagnosis per se have an impact on the performance of figurative language tasks. Indeed, the results of the meta-regression analysis revealed that the group matching strategy might explain some of the between-study variance. A further examination of the matching strategy showed differences in effect size in terms of whether the ASD and TD groups were matched based on CA, LA or both. In particular, the mean effect size was small (Hedges' $g=-0.06$) for studies that used LA as a matching variable, whereas the effect size was large (Hedges' $g=-0.92$) for studies that included groups that were matched according to CA. Thus, when individuals with ASD and TD individuals are compared based on their performance on core language tests, the former and the latter receive comparable scores on figurative language tasks. This means that figurative language comprehension in individuals with ASD is closely related to their core language skulls.

When interpreting these findings, it is important to note that several studies with CA-matched groups revealed that the ASD and TD groups exhibited significant inter-group differences in language skills (e.g. Rundblad and Annaz, 2010b). These differences would have influenced these groups' performances on figurative language tasks. In addition, some studies that used LA as a group matching variable showed at least slight group difference in CA. Namely, the participants with ASD were at least slightly older than the TD controls. In these studies, the high performance of participants with ASD on the figurative language tasks may well reflect the impact of age. This finding is supported by the results of studies that used both CA- and LA-matched groups to examine the effect size. Namely, when individuals with ASD are compared with both CA- and LA-matched TD controls, the former show poorer figurative language comprehension.

However, this result may also be explained by the verbal ability tests used to match the ASD and TD groups. In particular, the most frequently used tests in studies that involved groups matched based on both CA and LA were the verbal subtests of Wechsler intelligence tests. Only a few studies used other tests such as the Peabody Picture Vocabulary Test (PPVT) or the British Picture Vocabulary Scale (BPVS), the Test of Word Knowledge (TOWK), and the National Adult Reading Test (Nelson and Wilson, 1991). Whereas the studies that included LA-matched groups used BPVS, the Concepts and Directions Subtest of the CELF-III (Semel et al., 2000), the Recalling Sentences Subtest of the CELF-III (Conti-Ramsden et al., 2001), the picture vocabulary subscale of the TOLD-P:3 (Newcomer and Hammill, 1997) and the Syntax Construction subtest of the CASL as a measure of LA to match the ASD and the TD groups.

Gernsbacher and Pripas-Kapit (2012) argue that verbal IQ (VIQ) scores may overestimate the language abilities of individuals with ASD and, in turn, obscure the language impairment of many individuals with ASD. Therefore, testing vocabulary alone is important but insufficient when assessing the language abilities of individuals with ASD. Consequently, when individuals with ASD are matched to TD controls based on VIQ scores only, a poor understanding of figurative language may be a result of problems with language comprehension that may not be easily detected by the VIQ.

The narrative summary of the results of the studies that included dependent subgroups also supports the findings with respect to the close relationship between core language skills and figurative language comprehension. Namely, when the same group of individuals with ASD is compared with both the CA- and LA-matched groups of TD individuals, the group differences are evident. In particular, the group differences between the ASD group and LA-matched TD group are very small or almost nonexistent, indicating that language skills – not an ASD diagnosis per se – are related to problems with figurative language comprehension by individuals with ASD.

The finding that a subset of individuals with ASD, who have comparable language skills to those of TD individuals, do not have problems with figurative language comprehension indicates that figurative language comprehension is seemingly not universal or unique to individuals with ASD. This finding is supported by studies that report deficits in figurative language comprehension over a broad range of disorders, including learning disabilities, aphasia, Alzheimer's disease and Williams syndrome (e.g. Lee and Kamhi, 1990; Papagno and Caporali, 2007; Rapp and Wild, 2011).

As ToM was not included as a predictor variable in the meta-regression, we cannot make any claims about the independent contribution of ToM to figurative language comprehension. It is important to note that the finding with respect to the close relationships between core language skills and figurative language does not mean that ToM is not required to understand figurative language. Given the close relationship between core language skills and ToM, we can posit that high scores on core language tests may also be related to high scores on ToM tasks.

Our narrative summary has revealed that the findings of studies that examine the relationship between ToM and figurative language comprehension in individuals with ASD are inconsistent. On the one hand, some studies find that, once language is considered, ToM, as measured by false-belief tasks, does not explain the unique variance in, for example, metaphor comprehension (e.g. Norbury, 2004, 2005). On the other hand, a study by Whyte et al. (2014) concluded that ToM, as measured by, for example, 'Strange Stories' (Happé, 1994; O'Hare et al., 2009) and the children's version of the Reading the Mind in the Eyes (RMTE) task (Baron-Cohen et al., 2001), makes a unique contribution to figurative language (i.e. idioms) comprehension, even when core language skills are considered.

The use of different ToM measures may explain the varying results. For example, false-belief tasks are known to strongly depend on basic aspects of language that may be delayed or impaired in individuals with ASD (Hale and Tager-Flusberg, 2003; Milligan et al., 2007). In addition, false-belief tasks may not be sensitive to the continued development of ToM into adolescence (Wellman et al., 2001). The 'Strange Stories' (Happé, 1994; O'Hare et al., 2009) and the children's version of the RMTE task (Baron-Cohen et al., 2001) are considered to be more developmentally sensitive than false-belief tasks (Whyte et al., 2014).

However, the 'Strange Stories' are also closely related to verbal abilities. Teasing out the effects of cognitive and linguistic factors on figurative language comprehension is difficult (Norbury, 2005), partly because ToM tasks are verbally loaded and language plays an important role in the development of ToM (Lohmann and Tomasello, 2003; Nilsson and de López, 2016).

Is metaphor comprehension more difficult than irony and sarcasm comprehension for individuals with ASD relative to TD individuals?

In line with the theoretical literature on figurative language in TD individuals that claims that different tropes, for example, metaphor and irony, involve different pragmatic mechanisms and are thus understood differently by TD individuals (e.g. Wilson and Sperber, 2012), our study found that individuals with ASD also comprehend different types of figurative language differently. However, in contrast to, for example, Happé's (1993) study, we found that, compared with TD controls, individuals with ASD found metaphor comprehension more difficult than irony and sarcasm comprehension. One potential explanation for this unexpected finding may be the methodological characteristics of the studies, particularly the differences in the tasks used to measure different aspects of the same tropes of figurative language comprehension.

Research involving TD individuals has continually shown that an understanding of metaphor critically hinges on the task's complexity and on the effort that a person needs to make to complete this task. For example, metaphor comprehension tasks that require metalinguistic abilities (verbal definition tasks) are more demanding than, for example, multiple-choice tasks (Pouscoulous, 2014). However, multiple-choice tasks have also been questioned with respect to their ecological validity because they include literal interpretations that make them either too simplistic or too confusing (Norbury, 2004). The studies included in this meta-analysis used both multiple-choice (e.g. Adachi et al., 2004) and verbal definition tasks (e.g. Norbury, 2004; Whyte et al., 2014). In addition, many metaphor studies did not differentiate between novelty and conventionality, which are variables that may reliably explain the variance in performance.

Irony and sarcasm comprehension studies also used different tasks and different ways of administering these tasks. Although individuals with ASD are often reported to perform poorly on these tasks, some studies have found that these individuals can comprehend verbal irony in the context of computer-mediated tasks, which impose minimal social and verbal demands (e.g. Glenwright and Agbayewa, 2012). Several of the included studies used irony/sarcasm tests without too many verbal demands, which may explain the less pronounced differences between ASD and TD groups.

Are age and cross-linguistic differences less important?

The results of our study revealed that the group difference was stable across the age range studied. However, this finding should be interpreted with caution because some earlier studies involved, for example, a wide age range of participants, spanning from childhood through late adolescence (e.g. Landa and Goldberg, 2005). Another potential explanation for this finding may be that, as claimed by Pouscoulous (2011), children's linguistic competence, world knowledge and cultural expertise may account for the improvement in their figurative language abilities with age – not age per se. Thus, the association between language and figurative language does not seem to merely reflect the influence of age.

The size of the gap between the ASD and TD groups was the same in studies conducted in English and in those conducted in other languages, which suggests that, in the included studies, seemingly no cross-linguistic differences exist with regard to the figurative language comprehension of individuals with ASD. This finding conflicts with some findings in the cognitive linguistic literature, which discusses the variation of metaphor across languages and cultures in a TD context. In particular, evidence has shown that comprehension of some aspects of, for example, metaphor may vary between languages (e.g. Özçalışkan, 2003). More studies in different languages will be beneficial to understand figurative language comprehension in individuals with ASD from a cross-linguistic perspective.

Methodological issues

Some limitations observed in the included studies must be considered when interpreting the findings of this review. In a number of earlier studies, the main methodological shortcoming involved not performing appropriate tests to control for language comprehension in individuals with ASD. In addition, the wide age range, significant intergroup differences in language skills (e.g. Rundblad and Annaz, 2010b) and the lack of reading skill assessments when the task is administered in written form could have a considerable impact on the participants' performances on figurative language tasks.

How the studies operationalize the figurative language trope that they examine is often unclear, as are hypotheses based on theories of figurative language comprehension. This may drawback results in tasks that measure different aspects of figurative language, which require different pragmatic processes, thereby making it difficult to summarize the figurative language comprehension deficits in individuals with ASD.

However, in more recent studies, the methodology has improved. For instance, when examining figurative language comprehension, more studies use language measures that are more advanced than VIQ and vocabulary tests. Another example of recent methodological improvements involves including one ASD group and two TD control groups – one matched based on CA and another matched based on LA. This approach is useful in examining whether figurative LA is developmentally delayed or deviant in individuals with ASD.

Moreover, several studies used analysis of variance (ANOVA) and regression analyses (e.g. Norbury, 2004), and few studies used cross-sectional developmental trajectory analysis, which is based on developmental trajectories or growth models (Thomas et al., 2009). The latter is seemingly a well-justified alternative for group matching in ASD research because it does not require a priori group matching, which has been shown to be challenging for individuals with ASD (Jarrold and Brock, 2004).

Limitations of this meta-analysis

Meta-analyses are generally criticized for ignoring important between-study differences by creating a summary of the outcomes (Borenstein et al., 2009). However, the potential sources of between-study differences may be formally addressed by examining potential predictor variables. Unfortunately, one actual predictor that could have reliably explained between-study differences, ToM, could not be included in the meta-regression analysis because of the insufficient number of studies.

Collapsing multiple outcomes in the studies and creating a composite score to avoid threats related to using dependent data resulted in the following limitation: in the studies that included dependent subgroups (one target group and two control groups (e.g. Whyte et al., 2014) or one control group and two ASD groups (Norbury, 2004), we combined the outcome measures to ensure that dependent data contributed to one effect size only. The main point of the methodologies used in these studies was initially to show that, when a group of individuals with ASD is compared with an LA-matched TD group, the group difference is generally small and nonsignificant. By combining the outcomes into one composite score, these important differences are not highlighted in this meta-analysis. The meta-analysis instead asks whether figurative language deficits are specific to individuals with ASD relative to TD individuals in general. However, we addressed this issue by narratively summarizing and discussing the findings of the three studies that included dependent subgroups.

Implications for practice and research

Given that a subset of individuals with ASD display greater deficits compared with CA-matched TD controls in comprehending various figurative language tropes, these difficulties should be addressed in clinical and educational settings.

It is important that teachers, parents and clinicians do not avoid using figurative language in their interactions with children and adolescents with ASD, as figurative language frequently occurs in daily life settings. Instead, parents, teachers and clinicians should use and monitor figurative language with great awareness, provide explanations when required, and reinforce comprehension across settings (see also Kerbel and Grunwell, 1997, for similar recommendations with respect to idioms).

Because the deficits in figurative language appear to be related to core language skills, improving core language skills through educational and clinical interventions may lead to improved figurative language comprehension.

Given that controlling for language comprehension and vocabulary and VIQ measures is crucial to obtain valid results in figurative language comprehension studies on individuals with ASD (Gernsbacher and Pripas-Kapit, 2012), future studies on individuals with ASD should control for general language comprehension to eliminate the impact of this covariate when studying figurative language. Likewise, more studies are needed to investigate the role of syntax in figurative language comprehension in individuals with ASD.

More research is needed to examine the independent contributions of core language skills and ToM in figurative language comprehension. It is important that studies examining the role of ToM use developmentally sensitive and appropriate ToM tasks, preferably ones that are not highly correlated with verbal ability (Whyte and Nelson, 2015).

The causal relationships between, for example, ToM and LA and figurative language comprehension can be established by explicitly targeting figurative language skills through training studies. A few interventions that are designed for figurative language comprehension have been shown to be successful in children with ASD (e.g. Mashal and Kasirer, 2011; Persicke et al., 2012; Whyte et al., 2013). Figurative language skill training would also benefit adults and adolescents with ASD.

The quality of future studies should be optimized by considering the challenges related to small samples as well as those with wide age ranges. Cross-sectional developmental trajectory analysis is one possible approach to overcome the latter challenge (Thomas et al., 2009). However, the most informative way to study the development of figurative language comprehension, is to longitudinally examine the development of the same participants over time. Notably, no longitudinal studies were found in our systematic literature search.

Although we conclude that group matching strategy and trope differences could explain some of the variance in effect size, a large proportion of the variance has yet to be explained. Several critical variables, such as task content, task demands and task administration, deserve more careful consideration in future research.

Conclusion

In conclusion, individuals with ASD show deficits in figurative language comprehension compared with their TD controls. The significant differences between the studies can reliably be explained by trope differences, indicating that different tropes require different comprehension processes in individuals with ASD. Moreover, the group matching strategy was a reliable predictor of figurative language comprehension. Namely, the studies that matched groups based on LA yielded small and nonsignificant effect sizes, indicating that figurative language comprehension in individuals with ASD is closely related to core language abilities but simultaneously disputing the view of the uniqueness and universality of figurative language comprehension deficits in individuals with ASD. Therefore, interventions and educational programmes that aim to improve the social communication skills of individuals with ASD should target core language skills in addition to social skills.

Acknowledgements

The authors would like to thank the Librarian at the University of Oslo, Glenn Karlsen Bjerkenes, for assisting in literature search process, research assistant Håkon Næss for helping with screening the studies for eligibility, doctoral candidate Hanne Næss Hjetland for assisting in double-coding of the data, professor emeritus Anne Edwards, professor Hanne Gram Simonsen, Dr Alexander Tsertsvadze and researcher, PhD, Ingrid Lossius Falkum, for their useful comments and fruitful discussions with regard to the different aspects of the article. The authors would also thank the authors who provided the information that was necessary to conduct this meta-analysis.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

References

References marked with an asterisk denote studies included in the meta-analysis.

Ackerman BP (1982) On comprehending idioms: do children get the picture? *Journal of Experimental Child Psychology* 33(3): 439–454.

*Adachi T, Koeda T, Hirabayashi S, et al. (2004) The metaphor and sarcasm scenario test: a new instrument to help differentiate high functioning pervasive developmental disorder from attention deficit/hyperactivity disorder. *Brain & Development* 26(5): 301–306.

American Psychiatric Association (APA) (2013) *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th ed. Arlington, VA: APA.

*Au-Yeung SK, Kaakinen JK, Liversedge SP, et al. (2015) Processing of written irony in autism spectrum disorder: an eye-movement study. *Autism Research* 8: 749–760.

Baron-Cohen S (2001) Theory of mind in normal development and autism. *Prisme* 34: 174–183.

Baron-Cohen S, Leslie AM and Frith U (1985) Does the autistic child have a 'theory of mind'. *Cognition* 21(1): 37–46.

Baron-Cohen S, Wheelwright S, Hill J, et al. (2001) The 'reading the mind in the eyes' test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry* 42(2): 241–251.

Borenstein M, Hedges LV, Higgins JPT, et al. (2009) *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons.

Brynskov C, Krøjgaard P and Eigsti I (2016) Language and communication in children with autism: do research and clinical practice converge in Denmark? *Nordic Psychology* 68(2): 100–113.

Cain K, Oakhill J and Lemmon K (2005) The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology* 90(1): 65–87.

Carrow-Woolfolk E (1999) *Comprehensive Assessment of Spoken Language*. Circle Pines, MN: AGS Publishing.

*Channon S, Crawford S, Orlowska D, et al. (2014) Mentalising and social problem solving in adults with Asperger's syndrome. *Cognitive Neuropsychiatry* 19(2): 149–163.

*Chouinard B and Cummine J (2016) All the world's a stage: evaluation of two stages of metaphor comprehension in people with autism spectrum disorder. *Research in Autism Spectrum Disorders* 23: 107–121.

Cohen J (1968) Multiple regression as a general data-analytic system. *Psychological Bulletin* 70(6): 426–443.

*Colich NL, Wang A, Rudie JD, et al. (2012) Atypical neural processing of ironic and sincere remarks in children and adolescents with autism spectrum disorders. *Metaphor and Symbol* 27(1): 70–92.

Colston HL and Gibbs RW (2002) Are irony and metaphor understood differently? *Metaphor and Symbol* 17(1): 57–60.

Colston HL and Kuiper MS (2002) Figurative language development research and popular children's literature: why we should know, 'where the wild things are'. *Metaphor and Symbol* 17(1): 27–43.

Conti-Ramsden G, Botting N and Faragher B (2001) Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines* 42: 741–748.

*De Villiers PA, de Villiers JG, Diaz S, et al. (2011) Non-literal language and theory of mind in autism spectrum disorders. In: *Poster presented at the American speech-language hearing association convention*, San Diego, CA, 16–19 November.

*Dennis M, Lazenby AL and Lockyer L (2001) Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders* 31(1): 47–54.

Dews S, Winner E, Kaplan J, et al. (1996) Children's understanding of the meaning and functions of verbal irony. *Child Development* 67: 3071–3085.

Dunn LM, Dunn LM, Whetton C, et al. (1997) *The British Picture Vocabulary Scale*. 2nd ed. Windsor: NFER-Nelson.

Eigsti I, de Marchena AB, Schuh JM, et al. (2011) Language acquisition in autism spectrum disorders: a developmental review. *Research in Autism Spectrum Disorders* 5(2): 681–691.

Falkum IL, Recasens M and Clark EV (2016) 'The moustache sits down first': on the acquisition of metonymy. *Journal of Child Language* 19: 1–33.

Gernsbacher MA and Pripas-Kapit SR (2012) Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. *Metaphor and Symbol* 27(1): 93–105.

Gibbs RW (2000) Irony in talk among friends. *Metaphor and Symbol* 15(1–2): 5–27.
Gibbs RW and Colston HL (2012) *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.
*Glenwright M and Agbayewa AS (2012) Older children and adolescents with high-functioning autism spectrum disorders can comprehend verbal irony in computer-mediated communication. *Research in Autism Spectrum Disorders* 6(2): 628–638.
Gluksberg S (2001) *Understanding Figurative Language: From Metaphor to Idioms* (Oxford Psychology Series). New York: Oxford University Press.
*Gold R and Faust M (2010) Right hemisphere dysfunction and metaphor comprehension in young adults with Asperger syndrome. *Journal of Autism and Developmental Disorders* 40(7): 800–811.
*Gunter HL, Ghaziuddin M and Ellis HD (2002) Asperger syndrome: tests of right hemisphere functioning and interhemispheric communication. *Journal of Autism and Developmental Disorders* 32(4): 263–281.
Hancock JT, Dunham PJ and Purdy K (2000) Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development* 1(2): 227–248.
Happé FG (1993) Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48(2): 101–119.
Happé FG (1994) An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders* 24(2): 129–154.
Hedges LV (1981) Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics* 6(2): 107–128.
Hedges LV and Olkin I (1985) *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
*Hermann I, Haser V, van Elst LT, et al. (2013) Automatic metaphor processing in adults with Asperger syndrome: a metaphor interference effect task. *European Archives of Psychiatry and Clinical Neurosciences* 263 (Suppl. 2): S177–S187.
Hobson RP (2012) Autism, literal language and concrete thinking: some developmental considerations. *Metaphor and Symbol* 27(1): 4–21.
*Huang S, Oi M and Taguchi A (2015) Comprehension of figurative language in Taiwanese children with autism: the role of theory of mind and receptive vocabulary. *Clinical Linguistics & Phonetics* 29(8–10): 764–775.
*Imaizumi S, Furuya I and Yamasaki K (2009) Voice as a tool communicating intentions. *Logopedics, Phoniatrics, Vocology* 34(4): 196–199.
Jarrold C and Brock J (2004) To match or not to match? Methodological issues in autism-related research. *Journal of Autism and Developmental Disorders* 34(1): 81–86.
*Kaland N, Møller-Nielsen A, Callesen K, et al. (2002) A new 'advanced' test of theory of mind: evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology and Psychiatry* 43(4): 517–528.
*Kasirer A and Mashal N (2014) Verbal creativity in autism: comprehension and generation of metaphoric language in high-functioning autism spectrum disorder and typical development. *Frontiers in Human Neuroscience* 8: 615.
Kerbel D and Grunwell P (1997) Idioms in the classroom: an investigation of language unit and mainstream teachers' use of idioms. *Child Language Teaching & Therapy* 13(2): 113–123.
Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4: 863.
Lakoff G and Johnson M (2003) *Metaphors We Live by*. Chicago, IL: University of Chicago Press.
*Landa RJ and Goldberg MC (2005) Language, social, and executive functions in high functioning autism: a continuum of performance. *Journal of Autism and Developmental Disorders* 35(5): 557–573.
Lee RF and Kamhi AG (1990) Metaphoric competence in children with learning disabilities. *Journal of Learning Disabilities* 23(8): 476–482.
*Lee SB, Song SH, Ham JH, et al. (2015) Idiom comprehension deficits in high-functioning autism spectrum disorder using a Korean autism social language task. *Yonsei Medical Journal* 56(6): 1613–1618.
*Lewis FM, Murdoch BE and Woodyatt GC (2007) Communicative competence and metalinguistic ability: performance by children and adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 37(8): 1525–1538.
*Li JP, Law T, Lam GY, et al. (2013) Role of sentence-final particles and prosody in irony comprehension in Cantonese-speaking children with and without autism spectrum disorders. *Clinical Linguistics & Phonetics* 27(1): 18–32.
Lyons V and Fitzgerald M (2004) Humor in autism and Asperger syndrome. *Journal of Autism and Developmental Disorders* 34(5): 521–531.
*McCrimmon AW, Schwean VL, Saklofske DH, et al. (2012) Executive functions in Asperger's syndrome: an empirical investigation of verbal and nonverbal skills. *Research in Autism Spectrum Disorders* 6: 224–233.
*MacKay G and Shaw A (2004) A comparative study of figurative language in children with autistic spectrum disorders. *Child Language Teaching & Therapy* 20(1): 13–32.
*Martin I and McDonald S (2004) An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *Journal of Autism and Developmental Disorders* 34(3): 311–328.
*Mashal N and Kasirer A (2011) Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities* 32(6): 2045–2054.
*Mathersul D, McDonald S and Rushby JA (2013) Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology* 35(6): 655–668.
Melogno S, D'Ardia C, Pinto MA, et al. (2012) Explaining metaphors in high-functioning autism spectrum disorder children: a brief report. *Research in Autism Spectrum Disorders* 6(2): 683–689.
*Minshew NJ, Goldstein G and Siegel DJ (1995) Speech and language in high-functioning autistic individuals. *Neuropsychology* 9(2): 255–261.

Næss KA, Lyster SA, Hulme C, et al. (2011) Language and verbal short-term memory skills in children with down syndrome: a meta-analytic review. *Research in Developmental Disabilities* 32(6): 2225–2234.

Nelson HE and Wilson JR (1991) *The Revised National Adult Reading Test*. Windsor: NFER-Nelson.

Newcomer PL and Hammill DD (1997) *Test of Language Development, Primary*. 3rd ed. Austin, TX: Pro-Ed.

Nippold MA (1991) Evaluating and enhancing idiom comprehension in language-disordered students. *Language, Speech, and Hearing Services in Schools* 22(3): 100–106.

*Norbury CF (2004) Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research* 47(5): 1179–1193.

Norbury CF (2005) The relationship between theory of mind and metaphor: evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology* 23(3): 383–399.

*Olofson EL, Casey D, Oluyedun OA, et al. (2014) Youth with autism spectrum disorder comprehend lexicalized and novel primary conceptual metaphors. *Journal of Autism and Developmental Disorders* 44(10): 2568–2583.

Özçalışkan S (2003) Metaphorical motion in crosslinguistic perspective: a comparison of English and Turkish. *Metaphor and Symbol* 18(3): 189–228.

*Ozonoff S and Miller JN (1996) An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and Language* 52(3): 411–434.

Papagno C and Caporali A (2007) Testing idiom comprehension in aphasic patients: the effects of task and idiom type. *Brain and Language* 100(2): 208–220.

Passanisi A and Di Nuovo S (2015) Social and pragmatic impairments in individuals with autism spectrum disorder. A lack of theory of mind? *Life Span and Disability* 18(1): 75–99.

Persicke A, Tarbox J, Ranick J, et al. (2012) Establishing metaphorical reasoning in children with autism. *Research in Autism Spectrum Disorders* 6(2): 913–920.

*Peterson CC, Wellman HM and Slaughter V (2012) The mind behind the message: advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development* 83(2): 469–485.

Pexman PM (2008) It's fascinating research: the cognition of verbal irony. *Current Directions in Psychological Science* 17(4): 286–290.

Pexman PM and Glenwright M (2007) How do typically developing children grasp the meaning of verbal irony? *Journal of Neurolinguistics* 20(2): 178–196.

Pexman PM, Glenwright M, Krol A, et al. (2005) An acquired taste: children's perceptions of humor and teasing in verbal irony. *Discourse Processes* 40(3): 259–288.

*Pexman PM, Rostad KR, McMorris CA, et al. (2011) Processing of ironic language in children with high-functioning autism spectrum disorder. *Journal of Autism and Developmental Disorders* 41(8): 1097–1112.

Pons F, de Rosnay M, Harris P, et al. (2009) Theory-of-mind and language in children. *Impuls Tidsskrift for Psykologi* 3: 30–34.

Pouscoulous N (2011) Metaphor: for adults only? *Belgian Journal of Linguistics* 25: 51–79.

Pouscoulous N (2014) The elevator's buttocks metaphorical abilities in children. In: Matthews D (ed.) *Pragmatic Development in First Language Acquisition*. Amsterdam: John Benjamins, pp.239–259.

*Rajendran G, Mitchell P and Rickards H (2005) How do individuals with Asperger syndrome respond to nonliteral language and inappropriate requests in computer-mediated communication? *Journal of Autism and Developmental Disorders* 35(4): 429–443.

Rapp AM and Wild B (2011) Nonliteral language in Alzheimer dementia: a review. *Journal of the International Neuropsychological Society* 17(2): 207–218.

Rundblad G and Annaz D (2010a) Development of metaphor and metonymy comprehension: receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology* 28(3): 547–563.

*Rundblad G and Annaz D (2010b) The atypical development of metaphor and metonymy comprehension in children with autism. *Autism* 14(1): 29–46.

Saban-Bezalel R and Mashal N (2015) Hemispheric processing of idioms and irony in adults with and without pervasive developmental disorder. *Journal of Autism and Developmental Disorders* 45(11): 3496–3508.

Samson AC (2013) Humor(lessness) elucidated – sense of humor in individuals with autism spectrum disorders: review and introduction. *Humor* 26(3): 393–409.

Scammacca N, Roberts G and Stuebing KK (2014) Meta-analysis with complex research designs: dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research* 84(3): 328–364.

*Scheeren AM, de Rosnay M, Koot HM, et al. (2013) Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry* 54(6): 628–635.

Semel EM, Wiig EH and Secord W (2000) *Clinical Evaluation of Language Fundamentals*. 3rd ed. London: The Psychological Corporation.

Semrud-Clikeman M and Glass K (2010) The relation of humor and child development: social, adaptive, and emotional aspects. *Journal of Child Neurology* 25(10): 1248–1260.

Sperber D and Wilson D (1986) *Relevance: Communication and Cognition*. Oxford: Blackwell Publishing House.

*Strandburg RJ, Marsh JT, Brown WS, et al. (1993) Event-related potentials in high-functioning adult autistics: linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia* 31(5): 413–434.

Swineford LB, Thurm A, Baird G, et al. (2014) Social (pragmatic) communication disorder: a research review of this new DSM-5 diagnostic category. *Journal of Neurodevelopmental Disorders* 6(1): 41.

Tager-Flusberg H (2006) Defining language phenotypes in autism. *Clinical Neuroscience Research* 6(3–4): 219–224.

Tager-Flusberg H and Joseph RM (2003) Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358(1430): 303–314.

Thomas MS, Annaz D, Ansari D, et al. (2009) Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research* 52(2): 336–358.

Vulchanova M, Saldaña D, Chahboun S, et al. (2015) Figurative language processing in atypical populations: the ASD perspective. *Frontiers in Human Neuroscience* 9: 24.

*Wang AT, Lee SS, Sigman M, et al. (2006) Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain* 129(4): 932–943.

Whyte EM and Nelson KE (2015) Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders* 54: 2–14.

Whyte EM, Nelson KE and Khan KS (2013) Learning of idiomatic language expressions in a group intervention for children with autism. *Autism* 17(4): 449–464.

*Whyte EM, Nelson KE and Scherf KS (2014) Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research* 57(1): 120–130.

*Williams DL, Cherkassky VL, Mason RA, et al. (2013) Brain function differences in language processing in children and adults with autism. *Autism Research* 6(4): 288–302.

Wilson D and Sperber D (2012) *Meaning and Relevance*. Cambridge: Cambridge University Press.

Winner E (1988) *The Point of Words: Children's Understanding of Metaphor and Irony*. Cambridge, MA: Harvard University Press.

Winner E, Rosenstiel AK and Gardner H (1976) The development of metaphoric understanding. *Developmental Psychology* 12(4): 289–297.

*Wu C, Tseng L, An C, et al. (2014) Do individuals with autism lack a sense of humor? A study of humor comprehension, appreciation, and styles among high school students with autism. *Research in Autism Spectrum Disorders* 8(10): 1386–1393.

*Zalla T, Amsellem F, Chaste P, et al. (2014) Individuals with autism spectrum disorders do not use social stereotypes in irony comprehension. *PLoS ONE* 9(4): e95568.

*Zheng Q, Jia Z and Liang D (2015) Metaphor and metonymy comprehension in Chinese-speaking children with high-functioning autism. *Research in Autism Spectrum Disorders* 10: 51–68.

Appendix 1. Characteristics of the included studies.

Study	N	Mean CA	Trope	Effect size (g)	95% CI	Equating strategy
Author (year)	ASD (TD)	ASD (TD)				
*Adachi et al. (2004)	54 (199)	118 (120)	Metaphor; Sarcasm (Combined)	–0.45	[–0.75, –0.15]	No difference in CA and VIQ
*Au-Yeung et al. (2015)	22 (20)	389 (286)	Irony	–0.51	[–1.11, 0.09]	Matched based on VIQ
*Channon et al. (2014)	21 (21)	480 (524)	Sarcasm	–0.83	[–1.45, –0.21]	Matched based on CA
*Chouinard and Cummine (2016)	13 (12)	401 (396)	Metaphor	–0.45	[–1.21, 0.31]	Matched based on CA and semantic knowledge
*Colich et al. (2012)	15 (15)	171 (158)	Irony	0.55	[–0.15, 1.26]	Matched based on CA and VIQ
Dennis et al. (2001)	8 (8)	119 (113)	Metaphor/idiom	–1.37	[–2.42, –0.33]	Matched based on CA
*De Villiers et al. (2011)	30 (28)	149 (151)	Irony Metaphor (Combined)	–0.82	[–1.35, –0.29]	Matched based on CA and VIQ
*Glenwright and Agbayewa (2012)	14 (14)	148 (139)	Ironic criticism	0.14	[–0.58, 0.86]	Matched based on verbal mental age and CA
*Gold and Faust (2010)	27 (36)	275 (296)	LVF/RH Conventional metaphor LVF/RH Novel metaphor RVH/LH Conventional metaphor RVH/LH Novel metaphor (Combined)	–0.52	[–1.02, –0.01]	Matched based on CA and VIQ
*Gunter et al. (2002)	8 (8)	195 (203)	Humour Written metaphor Novel metaphor (Combined)	–1.35	[–2.40, –0.30]	Matched based on CA and VIQ
*Hermann et al. (2013)	20 (20)	509 (421)	Metaphor	–0.38	[–1.00, 0.22]	Matched based on vocabulary
*Huang et al. (2015)	50 (50)	122 (127)	Irony Metaphor Sarcasm (Combined)	–0.52	[–0.91, –0.12]	Matched based on CA and vocabulary
*Imaizumi et al. (2009)	20 (24)	119 (114)	Sarcasm	–1.85	[–2.55, –1.15]	Matched based on CA
*Kaland et al. (2002)	21 (20)	189 (186)	Figure of speech Irony (Combined)	–1.26	[–2.27, –0.26]	Matched based on CA
*Kasirer and Mashal (2014)	17 (17)	253 (273)	Conventional metaphor Novel metaphor (Combined)	–0.47	[–1.14, 0.20]	Matched based on CA
Landa and Goldberg (2005)	19 (19)	132 (132)	Metaphoric expressions and figures of speech	–1.01	[–1.68, –0.35]	Matched based on CA and VIQ
Lee et al. (2015)	16 (10)	111,72 (111,6)	Matched idiom task Mismatched idiom task (Combined)	–1.04	[–1.86, –0.23]	Matched based on age and IQ
Lewis et al. (2007) (Adults)	17 (13)	418 (416)	Figurative language	–1.09	[–1.85, –0.33]	Matched based on CA
Lewis et al. (2007) (Children)	20 (18)	139 (138)	Figurative language	0.32	[–0.30, 0.95]	Matched based on CA
*Li et al. (2013)	13 (13) 12 (12)	125 (125)	Irony belief Irony intention (Combined)	0.16	[–0.64, 0.84]	Matched based on LA and CA

(Continued)

Appendix 1. (Continued)

Study	N	Mean CA	Trope	Effect size (*g*)	95% CI	Equating strategy
Author (year)	ASD (TD)	ASD (TD)				
MacKay and Shaw (2004)	19 (21)	116 (123)	Hyperbole; indirect request; irony; metonymy; rhetorical questions; understatement (Combined)	–0.95	[–1.66, –0.23]	Matched based on CA and LA
*Martin and McDonald (2004)	14 (24)	236 (237)	Irony	–1.09	[–1.78, –0.40]	Matched based on CA
Mashal and Kasirer (2011)	20 (20)	156	Idiom Metaphor (Combined)	–0.73	[–1.36, –0.10]	Matched based on CA and LA
*Mathersul et al. (2013)	40 (33)	446 (500)	Sarcasm	–0.67	[–1.14, –0.20]	Matched based on CA and vocabulary
McCrimmon et al. (2012) (Cluster A)	24 (12)	222 (222)	Proverb	0.66	[–0.02, 1.36]	Matched based on CA and VIQ
McCrimmon et al. (2012) (Cluster B)	9 (21)	222 (222)	Proverb	1.25	[0.42, 2.07]	Matched based on CA
*Minshew et al. (1995)	62 (50)	213 (203)	Metaphoric expressions	–1.26	[–1.66, –0.85]	Matched based on CA and VIQ
Norbury (2004)			Idiom TD vs ASL Idiom TD vs ASO (Combined)	–1.16	[–1.70, –0.62]	Matched based on CA
*Olofson et al. (2014)	13 (13)	155 (153)	Conventional metaphor Novel metaphor (Combined)	–0.90	[–1.69, –0.12]	Matched based on CA
Ozonoff and Miller (1996)	17 (17)	314 (287)	Humour	–0.91	[–1.60, –0.22]	Matched based on CA and VIQ
*Pexman et al. (2011)	18 (18)	132 (132) 132 (94)	Ironic compliment ASD vs TD CAM Ironic compliment ASD vs TD LAM Ironic criticism ASD vs TD CAM Ironic criticism ASD vs TD LAM (Combined)	0.05	[–0.59, 0.69]	Matched based on LA
*Peterson et al. (2012)	44 (29)	108 (105)	Sarcasm	–0.39	[–0.86, 0.07]	Matched based on CA
*Rundblad and Annaz (2010b)	11 (17)	101 (100)	Metaphor Metonymy (Combined)	–1.32	[–2.13, –0.51]	Matched based on CA. Significant group difference in verbal skills
*Rajendran et al. (2005)	9 (12)	198 (201)	Figure of speech Sarcasm (Combined)	–0.30	[–1.14, 0.52]	Matched based on CA and VIQ
Saban-Bezalel and Mashal (2015)	23 (24)	316 (327)	Idiom Irony (Combined)	–0.65	[–1.23, 0.07]	Matched based on CA and VIQ
*Scheeren et al. (2013) (Adolescents) *Scheeren et al. (2013) (Children)	84 (16) 19 (7)	184 (172) 122 (114)	Sarcasm Sarcasm	0.30 0.16	[–0.11, 0.72] [–0.39, 0.72]	Matched based on receptive IQ (significantly older ASD group)
Strandburg et al. (1993)	13 (13)	299 (314)	Idiom	–1.12	[–1.92, –0.31]	Matched based on CA and VIQ

Appendix 1. (Continued)

Study Author (year)	N ASD (TD)	Mean CA ASD (TD)	Trope	Effect size (g)	95% CI	Equating strategy
Wang et al. (2006)	18 (18)	143 (143)	Idiom	–0.88	[–1.56, –0.21]	Matched based on CA and VIQ
Whyte et al. (2014)		116 (114) 116 (114)	Idiom ASD vs CAM Idiom ASD vs LAM (Combined)	0.50	[–1.04, 0.04]	Matched based on CA and LA
*Williams et al. (2013) (Adults) *Williams et al. (2013) (Children)	13 (12) 15 (14)	299 (252) 156 (150)	Irony Irony	–0.91 –0.48	[–1.71, –0.11] [1.20, 0.23]	Matched based on CA and VIQ Matched based on CA and VIQ
Wu et al. (2014)	164 (164)	165 (165)	Incongruity Nonsense joke (Combined)	–0.52	[–0.74, –0.30]	Matched based on CA and IQ
*Zalla et al. (2014)	17 (17)	328 (361)	Irony	–0.58	[–1.25, 0.08]	Matched based on CA and VIQ
*Zheng et al. (2015)	15 (15)	78 (75)	Conventional metaphor Conventional metonyms Novel metaphors Novel metonyms (Combined)	–0.75	[–1.48, –0.03]	Matched based on CA and VIQ

ASD: autism spectrum disorder; ASL: ASD with language impairment; ASO: ASD only; CA: chronological age; CI: confidence interval; LA: language ability; LH: left hemisphere; LVF: left visual field; RH: right hemisphere; RVF: right visual field; TD: typically developing.
References marked with an asterisk denote studies included in subgroup analysis of metaphor and irony and sarcasm.