

# Is verbal reference impaired in autism spectrum disorder? A systematic review

Louise Malkin, Kirsten Abbot-Smith and David Williams

University of Kent, UK

Autism & Developmental Language Impairments

Volume 3: 1–24

© The Author(s) 2018

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/2396941518763166

journals.sagepub.com/home/dli



## Abstract

**Background and aims:** Pragmatic language is a key difficulty in autism spectrum disorder. One such pragmatic skill is verbal reference, which allows the current entity of shared interest between speakers to be identified and thus enables fluid conversation. The aim of this review was to determine the extent to which studies have found that verbal reference is impaired in autism spectrum disorder. We organise the review in terms of the methodology used and the modality (production versus comprehension) in which proficiency with verbal reference was assessed. Evidence for the potential cognitive underpinnings of these skills is also reviewed.

**Main contribution and methods:** To our knowledge, this is the first systematic review of verbal reference in autism spectrum disorder. PsychINFO and Web of Science were systematically screened using the combination of search terms outlined in this paper. Twenty-four studies met our inclusion criteria. Twenty-two of these examined production, whereby the methodology ranged from elicited conversation through to elicited narrative, the ‘director’ task and other referential communication paradigms. Three studies examined reference interpretation. (One study investigated both production and appropriacy judgement). Four studies examined the relationship between appropriate usage of verbal reference and formal language (lexico-syntactic ability). Two studies investigated whether reference production related to Theory of Mind or Executive Functioning.

**Conclusion and implications:** Across a range of elicited production tasks, the predominant finding was that children and adults with autism spectrum disorder demonstrate a deficit in the production of appropriate verbal reference in comparison not only to typically developing groups, but also to groups with Developmental Language Disorder or Down syndrome. In contrast, the studies of reference interpretation which compared performance to typical control groups all found no between-group differences in this regard. To understand this cross-modality discrepancy, we need studies with the same sample of individuals, whereby the task requirements for comprehension and production are as closely matched as possible. The field also requires the development of experimental manipulations which allow us to pinpoint precisely if and how each comprehension and/or production task requires mentalising and/or various components of executive functioning. Only through such detailed and controlled experimental work would it be possible to determine the precise location of impairments in verbal reference in autism spectrum disorder. A better understanding of this would contribute to the development of interventions.

## Keywords

Autism spectrum disorders, pragmatics, referential communication, narrative, audience design

Autism spectrum disorder (ASD) is defined by persistent deficits in social-communication alongside restricted, repetitive patterns of behaviour, interests or activities (American Psychiatric Association, 2013; *Diagnostic and Statistical Manual of Mental*

*Disorders, 5th ed* [DSM-5]). Social use of language, or ‘pragmatics’ (social verbal communication), is considered a central impairment in ASD (Landa, 2000). Difficulties in this area hinder the ability to establish and maintain reciprocity in conversation and impair

## Corresponding author:

Kirsten Abbot-Smith, Kent Child Development Unit School of Psychology, Keynes College, University of Kent, Canterbury CT2 7NP, UK.

Email: K.Abbot-Smith@kent.ac.uk



the successful exchange of relevant information necessary for collaboration, negotiation and daily interaction (Tager-Flusberg, Paul, & Lord, 2005). A deficit in pragmatics has additionally been linked to mental health difficulties in a number of populations (e.g. Helland, Lundervold, Heimann, & Posserud, 2014). Whether all domains of pragmatics are equally impaired or indeed impaired in all individuals with ASD, however, remains unclear (Simmons, Paul, & Volkman, 2014).

A core component of pragmatics is reference, that is the ability to denote an entity, person or event with sufficient clarity for one's interlocutor. While theorists from a semiotic perspective give equal weight to verbal and non-verbal means (see Perkins, 2005, for discussion), in the current review, we follow Norbury (2014) among others in distinguishing pragmatic language from social communication more broadly and therefore we focus solely on 'verbal reference'. Impaired verbal reference is highly likely to have a severe detrimental effect on conversational flow (and thus on social relationships) and on the ability to collaborate with others (e.g. Murphy, Faulkner, & Farley, 2014), which would also have educational consequences. While the appropriate production and interpretation of verbal reference is often included in speech and language assessment and intervention for individuals with ASD (e.g. Adams, Gaile, Freed, & Lockton, 2010), the extent to which this area of pragmatic ability is universally impaired in this population has not been examined to date.

Forms of verbal reference vary greatly in complexity, as presented in Table 1, from highly specific referring expressions such as 'the red ball on top of the cupboard' to pronominal referring expressions, such as 'she' or 'it'. Bare noun phrases (definite: 'the ball'; indefinite 'a ball') may be considered intermediate in terms of specificity (e.g. Gundel, Hedberg, & Zacharski, 1993). Correct use and interpretation of verbal reference requires children to acquire the network of potential forms used in his or her language.

Pronominal (and equivalent) systems, for example, show great cross-linguistic variation in their level of acquisition difficulty, depending on whether forms vary regarding factors such as the gender, plurality or honorific status of the referent and whether case needs to be marked. Children learning Germanic and Romance languages usually acquire pronominal form by the mid-preschool years (e.g. Chiat, 1986; Girouard, Ricard, & Decarie, 1997). The form of complex referring expressions can include, for example, adjectival phrases ('the red ball') or modifying phrases such as prepositional phrases ('the ball on the shelf') or relative clauses ('the ball that Daddy bought you'). Syntactic forms used for complex referring expressions are generally mastered by the late preschool years (see, e.g. Berman & Slobin, 1994; Brandt, 2011; Kidd & Cameron-Faulkner, 2008).

An adult-like mastery of verbal reference not only requires acquisition of form, but also the ability to vary the level of complexity in accordance with context. Referring terms are therefore matched to the informational needs of a specific interlocutor. The appropriate use of verbal reference is often described by appealing to Grice's (1975) theory of communication, in particular the co-operative principle specifying the maxims of quantity and manner. These specify that a speaker should provide sufficient information for the listener to determine reference but also be concise (i.e. speakers should not be over-informative). Although many researchers have since pointed out major problems for the Gricean account (e.g. Gergely & Csibra, 2005; Horn, 1984; Moore, 2014; Sperber & Wilson, 1995, see also Levinson, 1989), the framework still provides a useful means of conceptualising the types of skills required to carry out and understand acts of reference.

Regardless of theoretical perspective, the match (or mismatch) between a particular form of verbal reference (e.g. pronominal, bare noun phrase or complex referring expression) and a particular context can be judged as 'correct' or 'incorrect'. Here, 'context' can

**Table 1.** Forms of verbal reference.

Forms of verbal reference		
Pronominal and related forms of verbal reference	Pronoun	He/she/it/they
	Zero form	Laura yawned and ___ fell asleep
Deixis		That/this one/that girl
	Simple/bare noun phrase	The/a girl
Complex referring expression	Common noun	Laura
	Proper noun	The tall girl
	With adjective	The girl on the bike
	With modifying phrase	The girl that you met yesterday.

include the information that specific interlocutors know one another to share; for example, if a child knows that his father is well acquainted with his friend Jamie, then it would be over-informative to use a complex referring such as ‘the Jamie that came to my birthday party’ every time the referent ‘Jamie’ is introduced into the conversation. ‘Context’ can also include whether there are competing referents in the visual context. To illustrate, if there is only one brush in the vicinity, then asking a listener to pass ‘the brush’ may be sufficient. In contrast, if more than one brush is present, then the speaker may need to specify ‘the brush with the brown handle’. Finally, the relevant ‘context’ would also include how recently a referent has been mentioned. That is, if a referent has just been mentioned in dialogue or narrative, the speaker can usually (depending on whether there are competing referents) reduce the specificity of the referring term further by using pronouns. In this case, speaker and listener can use their knowledge of the shared common ground to determine which referents are likely to be most salient (Sperber & Wilson, 1995) and/or activated in working memory, which is usually considered to be the component of short-term memory used to manipulate and update concurrently incoming material (see Baddeley & Hitch, 1974).

For the purpose of this review, we focus on whether referring terms are appropriately based on the informational needs of the listener, given ‘context’ as defined above. Thus, when we say ‘appropriate’, this is not a qualitative judgement; a participant who says, ‘Give me the duck’ in a context in which he or she can see that the addressee can see two ducks should arguably receive a score equating to ‘incorrect’ for this particular request. However, since even typical adult speakers do not perform at ceiling in these types of tasks (Keysar, 2007) we use the terms ‘appropriate’ versus ‘inappropriate’ throughout, to describe the match or mismatch between the form and context.

The first aim of the current paper was therefore to carry out a systematic review to determine whether an impairment in the appropriate usage/interpretation of verbal reference is a global feature of ASD (or whether verbal reference is only impaired in individuals with ASD with comorbid intellectual or formal language difficulties). To this end, our focus was not on whether individuals with ASD used the same *forms* of reference (e.g. whether they use the same proportion of pronouns within, for example, a conversation as do typically developing controls). Rather, our focus was on whether individuals with ASD are atypical in their understanding of the ‘fit’ between reference form and context.

If we found that some studies did not report an impairment in verbal reference in ASD, our second research goal was to investigate the extent to which

this might be due to either to the methodology used or to the modality in which proficiency with verbal reference was measured. Finally, we also wished to investigate whether studies including individuals with ASD provide evidence regarding the cognitive underpinnings of verbal reference ability.

To determine our key search terms, we first attempted to pinpoint the types of tasks typically used to assess verbal reference, that is naturalistic interaction, narrative or the ‘director task’/referential communication paradigm (see Graf & Davies, 2014, for a review). We also attempted to identify the key concepts most commonly associated with verbal reference in the literature. One such concept is that of ‘listener needs’ or ‘audience design’; as previously described, to be optimally informative, a referring term should provide sufficient information without being over-informative. This type of adaption to the informational needs of the listener is considered appropriate audience design (Clark & Murphy, 1982). Successful audience design may be achieved through consideration of the information listener and speaker share, or ‘common ground’ (Clark & Marshall, 1981).

### Criteria for current review

Systematic searches were conducted in two databases: PsychINFO and Web of Science for all dates up until March 2016. Our search terms were entered into the ‘keyword’ field as follows: (a) *autis\** AND *narrative*, (b) *autis\** AND *referen\** AND *communicat\**, (c) *autis\** AND *common ground*, *autis\** AND *audience design*, (d) *autis\** AND *listener needs* and (e) *autis\** AND *director task*. Given that these two search engines are imperfect, it is inevitable that this review will not be exhaustive. Indeed, we found and included one study which met our search engine criteria (Kuijper, Hartman, & Hendriks, 2015), but which was detected by neither search engine. Nonetheless, this review should constitute an accurate representation of literature in this topic to date.

An initial review of titles and abstracts excluded studies that were clearly not related to the key topics of interest, such as articles on literature or politics. The remaining full articles were then examined and our inclusion criteria were applied as follows. To be included the study was required to (1) include participants with a diagnosis of either ASD, Asperger syndrome or pervasive development disorder-not otherwise specified, (2) include a measure of the appropriacy of the match between verbal (lexico-syntactic) reference and context, (3) contain quantitative data which were analysed statistically and (4) to include a control group consisting of either (a) typically developing individuals, (b) individuals with Developmental

Language Disorder (DLD) (Specific Language Impairment) or (c) individuals with an impairment in non-verbal (performance) IQ. Without one of these control groups it is difficult to conclude whether or not individuals with ASD are impaired in referential communication. A study was additionally excluded if: (a) it was a case study with a single participant (due to issues of generalisability), (b) it was a training study which did not contain sufficiently detailed baseline data for conclusions regarding impairment in verbal reference to be drawn or (c) it exclusively examined non-verbal communication such as gesture, facial expression and eye contact or solely considered prosody (rather than lexico-syntactic form). Studies exclusively examining non-verbal communication were excluded because we were primarily interested in verbal reference. One reason for this is that when investigating the role of cognitive underpinnings, it is likely that the role of formal language, in particular, would play quite a different role in relation to non-verbal reference than in relation to verbal reference.

The total number of studies considered for inclusion and those excluded at each stage in the search process are shown in Figure 1. The 24 studies that met our inclusion criteria are listed in Tables 2 to 7. First, studies comparing the *production* of verbal reference in ASD and TD groups are summarised. Next, studies that compare *comprehension* of verbal reference in ASD and TD groups are summarised. Finally, evidence for the potential cognitive underpinnings related to successful verbal reference is reviewed.

## Production of referring expressions

We first review studies that have used methodologies which most closely map onto naturalistic usage of verbal reference in daily life. We then review studies that have measured appropriate verbal reference during more structured narrative tasks, and finally the most structured elicitation technique, the referential communication task.

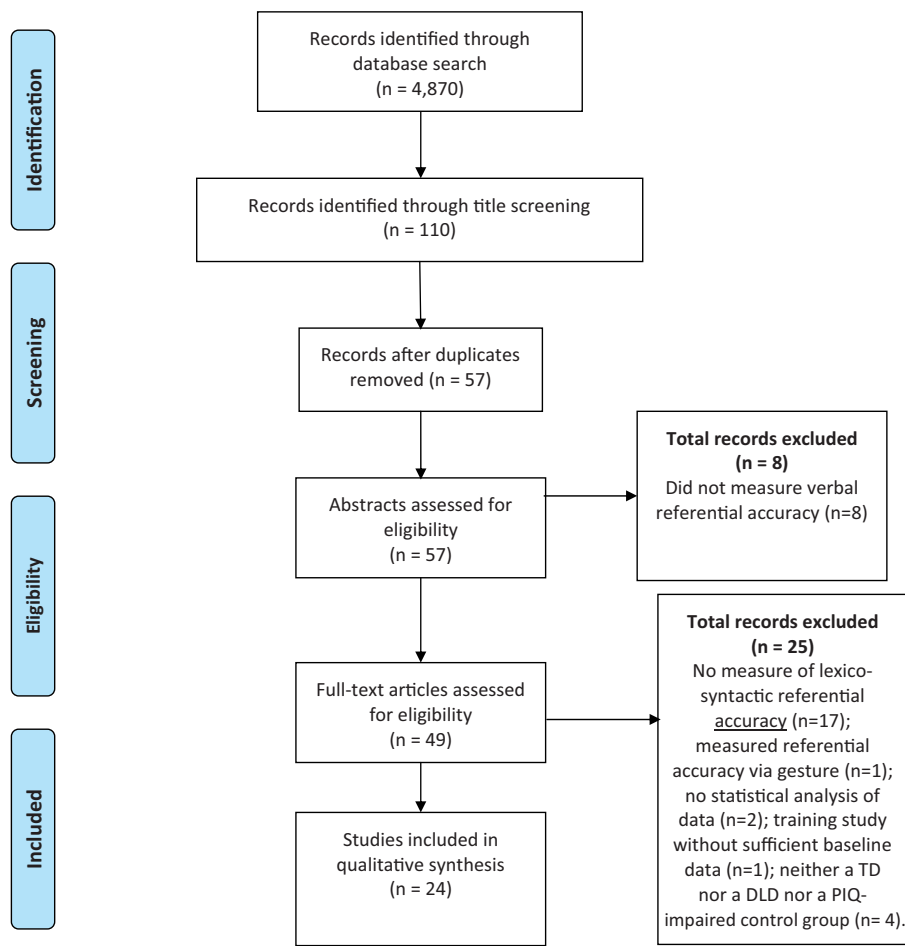


Figure 1. PRISMA diagram of study identification and selection.

### Production of referring terms during conversation

Though most closely mirroring real-life interaction, only one study containing at least one control group with clearly defined characteristics returned in our search measured the appropriacy of verbal reference use by individuals with ASD during conversation (see Table 2). In this study, Baltaxe and D'Angiola (1996) examined the use of pronominal, demonstrative (e.g. here/there) and comparative (e.g. bigger/smaller) reference during an hour-long interactive play session. Children with ASD ( $M = 7;9$  yrs,  $n=10$ ) were matched on language ability with a chronologically younger TD group ( $M = 3;5$  yrs,  $n=8$ ). Use of ambiguous reference (e.g. saying 'it' when the reference is unclear) was never found in the TD group (p. 252). This study also compared the ASD group to a group with DLD, ( $n=8$ ) matched on receptive language and Mean Length of Utterance. DLD is a diagnosis of language impairment in the absence of a known biomedical condition (Bishop, Snowling, Thompson, & Greenhalgh, 2016). In comparison to the DLD group, the ASD group used more ambiguous personal pronouns, though this failed to reach significance. This is presumably in part due to the extremely small sample size. Nonetheless, as we will see, the finding of a tendency towards ambiguity and the finding of deficits relative to children with DLD will be a recurrent theme throughout this review.

### Production of referring terms within narrative

**Overview.** Narrative tasks usually require the participant to generate or retell a story based on a picture book or film. They therefore constitute a monologue, rather than a reciprocal interaction. Therefore, narrative tasks might seem quite far removed from naturalistic verbal interaction. Nonetheless, narrative measures have been found to correlate strongly with standardised measures of pragmatic language more broadly, such as the Test of Pragmatic Language (e.g. Manolitsi & Botting, 2011).

Since we are interested in the degree to which individuals understand the *function* of verbal reference, our focus is on measures which assess whether the lexicosyntactic form is appropriate given the context. In the sentence 'Laura went to the shop and she bought some bread', for example, the initial reference 'Laura' is appropriate as it introduces a new character. The third-person subject pronoun 'she' is also appropriately unambiguous, referring the listener back to a character 'Laura' recently established as the focus of the conversation. The use of 'she' in this way is an example of '*anaphoric reference*'. Errors may be in the direction of over-informativity (e.g. if the full noun 'Laura' were

used throughout). When two potential referents have recently been mentioned, conversely, an anaphoric reference may be under-informative (e.g. 'I saw Laura and Karen and she looked upset').

Our survey of studies, which quantitatively measured referential accuracy within narrative, is organised in terms of the elicitation method employed. First, we review *narrative generation* studies, in which narrative is elicited from a stimulus (generally pictures depicting a story) without a prior model. Then, we review *narrative retell* studies, in which events are witnessed either in picture or video format and then retold either with reference to the original stimuli or from memory.

**Narrative generation.** We first discuss narrative generation studies where groups were not well matched (i.e. not matched for formal language ability, non-verbal IQ and chronological age), followed by those that were (see Table 3). The five less well-matched studies which examined narrative generation all included primary school-aged children (Colozzo, Morris, & Mirenda, 2015; Mäkinen et al., 2014; Norbury & Bishop, 2003; Norbury, Gemmell, & Paul, 2014; Tager-Flusberg, 1995). One of these studies (Norbury et al., 2014) also included young adolescents. Three of the five studies elicited narratives using the 24-page wordless story book 'Frog, where are you?' (Mayer, 1969) which has been used extensively to compare narrative generation cross-linguistically in TD children (see Berman & Slobin, 1994). One study found that children with ASD showed awareness of the rules for introducing and reintroducing characters in a narrative (Norbury & Bishop, 2003). That is, they were as likely as a TD group to use a noun phrase, rather than a pronoun, to reintroduce a character when it was not the most recently mentioned. All five studies, including Norbury and Bishop (2003), however, found evidence that children with ASD were more likely to use ambiguous reference than were typical controls.

Three of these studies also compared the ASD group to a group of individuals with DLD. Norbury and Bishop (2003) found that their ASD sample used more ambiguous nouns than did a DLD group matched for chronological age and language ability. The same pattern of results was found by Colozzo et al. (2015), whereby the ASD group possessed superior formal language ability than the DLD group but still used a higher number of ambiguous character references than did the DLD group. Finally, in Norbury et al. (2014), the difference between the ASD group and the DLD group did not reach significance. However, there was a moderate effect size ( $d=.47$ ) despite the fact that the ASD group in fact had significantly better formal language skills than the DLD group.

**Table 2.** Production – conversation/personal narrative.

Author, year	Participants			Matching		Measure of reference	RC impairment in ASD group		
	Sample size (n per group)	Mean age (range) per group	Language	ASD diagnosis	Structural Language			Age	Non-verbal IQ
Baltaxe and D'Angiola (1996)	n = 10 (ASD) n = 8 (TD) n = 8 (SLI)	ASD: 7;9 TD: 3;5 SLI: 7;7	English	Met DSM-III criteria for autism	✓ <u>MLU:ASD/TD:</u> d = 0.69 ASD/SLI: d = 0.14 <u>Receptive Vocabulary (PPVT):</u> ASD/TD: d = 0.19 ASD/SLI: d = 0.02 <u>Language Comprehension (TACL):</u> ASD/TD: d = 0.2 ASD/SLI: d = 0.29	X	– Not reported ASD/TD:d = -2.28 ASD/SLI:d = -0.03	Statistical analysis/inter-rater reliability Pronominal, demonstrative and comparative references, including ambiguous use. ANOVA % reliability	✓

ASD: autism spectrum disorder; MLU: mean length of utterance; ✓: well matched; PPVT: Peabody Picture Vocabulary Test; SLI: specific language impairment; TACL: Test for Auditory Comprehension of language; TD: typically developing; X: not well matched.

**Table 3.** Production – Narrative generation (stimulus present).

Author, year	Participants			Matching		Task	Measure of reference	RC impairment in ASD group		
	Sample size (n per group)	Mean age (range) per group	Language	ASD diagnosis	Formal language				Age	Non-verbal IQ
<b>Not well matched</b>										
Norbury and Bishop (2003)	n = 12 (ASD) n = 18 (TD) n = 17 (SLI) n = 21 (PLI)	ASD: 8.8 TD: 8.56 SLI: 9.33 PLI: 8.92 (all within 6–10 yrs)	English	SCQ/ADOS	Comprehension: BPVS: ASD/TD: d = 1.47 ASD/SLI: d = 0.40 TROG or CELF <sub>RLC</sub> : ASD/TD: d = 1.64 ASD/SLI: d = 0.52 Expressive Language CELF <sub>RS</sub> : ASD/TD: d = 2.83 ASD/SLI: d = 0.07	ASD/ITD d = 2.20 ASD/SLI d = 0.45	RCM ASD/TD d = 0.31 ASD/SLI d = 0.58	Frog where are you? ANOVA No IRR	Character reference (introduction, reintroduction and maintenance), including ambiguity	✓
Mäkinen et al. (2014)	n = 16	HFA: 7.7 (5–10) TD: 7.5 (5–10)	Finnish	ADOS/ADI-R	TTEC d = 1.25	✓ d = 0.11	NEPSY-II d = 1.21	The cat story t-tests/U Gd ICC	Character reference (including ambiguity)	×
Norbury et al. (2014)	n = 22	ASD: 11.1 (6.5–15.71) TD: 9.8 (6.7–15.1) LI: 10.7 (6.5–15.3)	English	Met criteria for ASD based on DSM-IV	CELF ASD/TD: d = 0.315	✓ ASD/TD: d = 0.47 ASD/SLI: d = 0.12	ASD/TD: d = 0.52 ASD/SLI: d = 0.22	A boy, a dog and a frog ANOVA % reliability	Character reference (including ambiguity)	✓
Colozzo et al. (2015)	n = 12	ASD: 8.5 (6–10) TD: 8.6 SLI: 8.5	English	Fulfilled diagnostic criteria for ASD according to experienced community-based clinicians	ASD/LI: d = 4.02 Only ASD performance reported	✓ ASD/TD: d = 0.07 ASD/SLI: d = 0.04	Only ASD performance reported	Late for School (wordless story)/ Aliens (single picture) from: Test of narrative language ANOVA % reliability	Character, object and place reference	Character ✓ Object × Place ×
<b>Well matched</b>										
Colle et al. (2008)	n = 12	ASD: 27.5 (n/r) TD: 27.2 (n/r)	English	ADI/ADOS	WAIS d = 0.01	✓ d = 0.02	WAIS d = 0.01	Frog where are you! Mann–Whitney U Cohen's k	Character (introduction, reintroduction, maintenance) and temporal references, including ambiguity	✓
Suh et al. (2014)	n = 15	ASD: 12.9 (10.5–15.7) TD: 13 (9.9–15.6)	English	ADOS	d = 0.66	✓ d = 0.16	d = 0.98	Tuesday ANOVA Gd ICC	Character reference (ambiguous pronouns)	✓
Banney et al. (2015)	n = 11 (ASD) n = 17 (TD)	ASD: 11.6 (9–15) TD: (9–15.2)	English	ADOS	CELF d = 0.63	✓ d = 0.27	RCM d = 0.37	Tuesday t-test/U % reliab.	Character reference (introduction, reintroduction and maintenance), including ambiguity	✓

(continued)

**Table 3. (continued)**

Author, year	Participants			Matching			Task	Measure of reference	RC impairment in ASD group	
	Sample size (n per group)	Mean age (range) per group	Language	ASD diagnosis	Formal language	Age				Non-verbal IQ
Kuijper et al. (2015)	n = 46 (ASD) n = 37 (ADHD) n = 38 (TD)	ASD: 9;3 (6;1-12;10) ADHD: 8;9 (6;1-12;10) TD: 9;0 (6;1-12;10)	Dutch	ADOS ADI-R	✓ PPVT ASD/TD d = 0.36 ASD/ADHD d = 0.28 × WISC Vocabulary ASD/TD d = 1.73 ASD/ADHD d = .30	✓ ASD/TD d = 0.14 ASD/ADHD d = 0.29	WISC Block Design ASD/TD d = 0.38 ASD/ADHD d = 0.45	4 stories of 6 pics developed by Hendriks, Koster and Hoeks (2014)	Specifically designed to measure appropriacy of character reference during (a) introduction, (b) maintenance (where no competing reference versus competing reference) and (c) reintroduction	×
<b>Matching not fully reported</b>										
Tager-Flusberg (1995)	n = 10	ASD: 12;1 TD: 7;9 MR: 11;3	English	Met DSM-III-R criteria for autism	✓ PPVT ASD/TD d = 0.56 ASD/MR d = 0.06	ASD/TD × ASD/MR ✓	– Not reported	Frog where are you? Kruskal-Wallis % relib.	Character reference (introduction and maintenance), including appropriateness of nouns/pronouns	✓ for character introduction × for 'referential style' n/s trend for TD to opt for more appropriate anaphoric style

ADI-R: Autism Diagnostic Interview-Revised; ADOS: Autism Diagnostic Observation Schedule; AS: Asperger syndrome; ASD: autism spectrum disorder; BPVS: British Picture Vocabulary Scale; CELF: Clinical Evaluation of Language Fundamentals; CELF<sub>RLC</sub>: Receptive Language Composite; CELF<sub>RS</sub>: recalling sentences; DSM: *Diagnostic and Statistical Manual of Mental Disorders*; HFA: high-functioning autism; HFA-F: female; HFA-M: male; ICC: intraclass correlation coefficient; ITPA: Illinois Test of Psychological Abilities; MLU: mean length of utterance; MR: mentally retarded; ✓: well matched; ×: not well matched.

RCM: Raven's coloured matrices; SCQ: Social Communication Questionnaire; SLI: specific language impairment; TD: typically developing; TD-F: female; TTFC: Token Test for Children; WAIS: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children; X: not well matched.



Only one of these five studies reported no differences between an ASD and TD group in production of ambiguous character reference (Mäkinen et al., 2014). This study was carried out with Finnish children aged 5–10 years. Groups were well matched for chronological age but the TD group scored higher for formal language and memory. One reason for the lack of a between-groups difference in this study might be that the participants were Finnish speaking and the authors note that in Finnish, TD children tend not to master accurate reference until eight years of age.

All of the above five narrative generation studies were not well matched to typical controls. Four additional studies examined narrative generation in comparison to well-matched controls. Three of these studies found impairments in individuals with ASD in comparison to typical controls. One of these studies focussed on adults and used the Mayer (1969) story ‘Frog, where are you?’ (Colle, Baron-Cohen, Wheelwright, & van der Lely, 2008), which was told to an experimenter who did not have visual access to the story pictures. Their sample with ASD used more ambiguous references to the dog and non-protagonist characters than did a TD group, despite being told that the listener had no previous knowledge of the story and that they should therefore be ‘as clear as possible’. The other three studies tested upper-primary school-aged children and adolescents. Two elicited narratives using a 29-page wordless picture book called ‘Tuesday’ by Wiesner (1991). These two studies reported that children and adolescents with ASD were more likely to use ambiguous reference in comparison to well-matched TD groups (Banney, Harper-Hill, & Arnott, 2015; Suh et al., 2014). The third study required participants to tell four stories to an experimenter, who did not have visual access to the pictures. Each story had two characters of the same gender and which were specifically constructed to examine reference selection for character introduction, character maintenance and character reintroduction (Kuijper et al., 2015). In this study, there were no significance differences in appropriacy of reference selection between the ASD and TD groups, despite large sample sizes (and despite the fact the ASD group scored significantly lower on the WISC ‘Vocabulary’ measure). However, the stories were much simpler than those used by the majority of narrative studies, both in terms of length (as each consisted solely of six pictures) and in terms of the amount of detail in each picture. This might have reduced both the working memory load of the task and the degree to which the individuals with ASD were likely to be distracted by irrelevant information. The potential issue of stimuli-dependent performance is one which will also emerge in the next section.

*Narrative retell.* In contrast to narrative *generation*, narrative *retell* instead involves listening to and/or viewing a story unfold, then retelling events from memory. Though not as widely used as generation tasks, four studies meeting our inclusion criteria utilised a retell method (see Table 4).

In the study by Arnold, Bennetto and Diehl (2009), children and adolescents with and without ASD watched a Sylvester and Tweety cartoon and then retold this from memory to a confederate who feigned ignorance of the story. There was no narrator dialogue in the video clip. Instead participants simply watched events unfold. Each character reference was coded for recency of mention of the antecedent. If a referent was mentioned no more than two clauses back, the *children* with ASD (9;8–12;9) used a significantly higher proportion of noun phrases (as opposed to pronouns) than did the typical controls, which the authors interpreted as over-informativity in this context. In contrast, the adolescents with ASD (13;1–17;8) did not differ from a well-matched TD group in any of the measures used. However, Arnold et al. (2009) did not assess the appropriacy per se of verbal reference selection; the latter does not solely depend on how many clauses back the antecedent was but rather, whether a referential alternative (e.g. Tweety Bird/Sylvester) was also recently mentioned and of course whether the pronouns are gender marked.

In addition, it is possible that the particular elicitation method/stimuli used partly accounts for discrepant findings between studies. To examine the extent to which elicitation method influences performance, Novogrodsky (2013) compared a narrative *retell* and *generation* task, analysing the ambiguity of third-person subject pronouns. The same data were reanalysed by Novogrodsky and Edelson (2016), whereby they extended their analysis to include subject and object pronouns. The retell task was the ‘Bus Story’ task (Renfrew, 1991), which requires the child to retell a story, which has first been told to the child, about a bus that escapes from its driver. In this task, participants can look at the pictures as they retell the story. In the generation task, children told ‘Frog, where are you?’ (Mayer, 1969) from pictures, without an initial model. Whilst ASD and TD group performance did not differ in terms of ambiguous pronominal reference during the retold Bus Story, in the generation task the ASD group used significantly more ambiguous pronouns.

Unfortunately, due to the design of this study, there are many potential reasons why results may have differed depending on the particular elicitation paradigm. First, since the children had just heard the administrator tell the Bus Story, those with good auditory recall (which is often a relative strength in ASD) might simply

**Table 4.** Production – narrative retell.

Author, year	Participants			Matching			Task	Measure of reference	RC impairment in ASD group	
	Sample size (n per group)	Mean age (range) per group	Language	ASD diagnosis	Structural language	Age				Non-verbal IQ
Arnold et al. (2009)	n = 10 (ASD-Y) n = 13 (ASD-O) n = 10 (TD-Y) n = 13 (TD-O)	ASD-Y: 11;6 (9;8–12;9) ASD-O: 15;1 (13;1–17;8) TD-Y: 11;6 (9;8–12;9) TD-O: 14;6 (13;1–17;8)	English	ADOS/ADI-R	VIQ: WISCWAIS ASD-Y/TD-Y d = 0.25 ASD-O/TD-O d = 0.20 Receptive Language: PPVT: ASD-Y/TD-Y d = 0.24 ASD-O/TD-O d = 0	✓ ASD-Y/TD-Y d = 0.53 ASD-O/TD-O d = 0.44	✓ WISCWAIS ASD-Y/TD-Y d = 0 ASD-O/TD-O d = 0.10	Sylvester and Tweety Cartoon Retell video (no audio) from memory	Statistical analysis/inter-rater reliability References to characters: noun (Sylvester; the cat) pronoun (he, it), zero (...and φ ran) Number of clauses since the most recent mention of the same character (1, 2, 3 clauses back, or no prior mention) ANOVA % reliability	Younger: ✓ Older: ×
Novogrodsky (2013)	n = 23 (ASD) n = 17 (TD)	ASD: 10 (6;1–14;3) TD: 9;9 (5;11–14;4)	English	ADOS/ADI-R	✓ Woodcock-Johnson III d = 0.30	✓ d = 0.11	Not reported	1. Retell: Bus Story (after having heard an adult tell story) 2. Generation: Frog where are you?	Third-person subject pronouns, including ambiguity	Retell × Generation ✓
Novogrodsky and Edelson (2016): reanalysing data from Novogrodsky (2013)	n = 24 (ASD) n = 17 (TD)	ASD: 10 (6;1–14;3) TD: 9;9 (5;11–14;4)	English	ADOS/ADI-R	✓ Woodcock-Johnson III d = 0.30	✓ SD not reported	Not reported	1. Retell: Bus Story 2. Generation: Frog where are you?	Use of subject, object and possessive pronouns, including ambiguity ANOVA IRR not reported	Retell × Generation (subject and possessive ambiguity) ✓ (object) × n.b. relatively low occurrence of object pronouns for both groups
de Marchena and Eigsti (2016)	n = 19	ASD: 14;10 (12;7–16;11) TD: 15;4 (12;2–17;11)	English	ADOS/SCQ/SRS	✓ Receptive Vocabulary PPVT p = .09 (marginal) d = 0.55 Expressive: Stanford-Binet Verbal d = 0	✓ d = 0.35	✓ Stanford-Binet NV d = 0.49	Cartoon clips Retell video (no audio) from memory	Referential shortening ANOVA Excellent IRR	✓

ADI-R: Autism Diagnostic Interview-Revised; ADOS: Autism Diagnostic Observation Schedule; ASD: autism spectrum disorder; ASD-O: older; ASD-Y: younger; MLU: mean length of utterance; ✓: well matched; PPVT: Peabody Picture Vocabulary Test; SCQ: Social Communication Questionnaire; SRS: Social Responsiveness Scale; TD-O: older; TD-Y: younger; VIQ: verbal IQ; WAIS: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children; X: not well matched.

have been able to select appropriate forms of verbal reference by recalling this ad verbum. Second, the ‘Bus Story’ only consists of 12 pictures and thus it could be that the relative simplicity of the story allowed more accurate use of reference.

The fourth study of narrative retell was conducted by de Marchena and Eigsti (2016), who used 60 second cartoon clips. This study differs from the other narrative studies in that listener informational needs were specifically manipulated by having two within-subjects conditions: ‘shared’ (the listener watched a short preview of the clip with the participant) and ‘private’ (the listener was not present during any part of the clip). Some aspects of de Marchena and Eigsti’s (2016) data indicate that the adolescents with ASD considered listener information needs to a degree; there was a significant difference in communicative quality ratings (i.e. a rating of how easy the story was to follow) between narratives produced by the ASD group in the shared versus the private condition. However, for the key dependent variable, the degree of referential shortening, there was a between-groups difference. The authors argue that the referential shortening effect is a measure of whether participants take audience needs into account. The argument is that, if speakers take audience needs into account, their narratives should be shorter when retelling in the shared as opposed to the private condition. This effect was seen for the typical control but not the ASD group, indicating that the latter had difficulty adapting to listener information needs. However, de Marchena and Eigsti’s analysis rests on the assumption that a longer narrative would contain a greater number of full noun phrases or indeed noun phrases with modifying phrases (see Table 1). This is of course not necessarily the case since a proper noun (e.g. Laura) is usually highly informative and yet does not differ in word length from a pronoun. Conversely, not all modifying phrases provide sufficient differentiating information. The extent to which reference selection was appropriate for a given context was not examined.

Nonetheless, there was a significant relationship between the referential shortening effect and symptom severity as measured by the Social Responsiveness Scale (Constantino & Gruber, 2007) in the ASD group, whereby those more likely to demonstrate the effect showed less ASD traits. This supports the authors’ conclusion that the referential shortening effect taps some of the social communicative deficits which are diagnostic for ASD. Older children with ASD were also more likely to show the referential shortening effect than those who were younger, tying in with Arnold et al.’s (2009) finding that selection of appropriate reference may improve with age in the ASD population.

*Referential communication tasks.* Over all narrative elicitation methods, the overwhelming tendency indicates an impairment in the appropriate usage (production) of verbal reference. However, it might be argued that the difficulties individuals with ASD experience with narrative tasks are not related to deficits in the production of appropriate verbal reference per se, but instead are related to extraneous demands required by these tasks. Individuals with ASD may be particularly hindered in narrative tasks by the need for episodic memory (e.g. Lind, Williams, Bowler, & Peel, 2014), executive functioning (e.g. Geurts, Verté, Oosterlaan, Roeyers, & Sergeant, 2004), imagination (e.g. Lind et al., 2014) and central coherence (e.g. Happé & Frith, 2006). Therefore, an elicitation method which does not burden episodic memory and which mirrors the back-and-forth nature of conversation might be better able to reveal underlying latent ability in individuals with ASD to use verbal reference appropriately. One such method is the referential communication task.

Referential communication tasks allow both the production and comprehension of referring terms to be measured. Here, we first review the results of studies where referential communication tasks have been used to examine the production of referring terms. Studies which examined the interpretation of reference are discussed later in this review. Our search returned five studies involving a type of referential communication paradigm to examine the appropriacy of referring terms selected in production.

A very frequently used type of referential communication task that allows visual common ground to be manipulated is the ‘Director Task’ (Keysar, Barr, Balin, & Brauner, 2000). Typically, an array of items or pictures is presented to a speaker, or ‘director’ who provides relevant information to a listener to enable them to select a target referent. The inclusion of items that differ in only one aspect, e.g. ‘*big* cup’ versus ‘*small* cup’ creates the need to alter the specificity of the referring terms selected. The *visual perspective* of the listener can be manipulated to either match or be discrepant with that of the speaker by blocking visual access to some referents in the array. Two potential conditions are therefore possible: A ‘shared’ condition, in which all potential referents can be viewed by both parties (i.e. visual perspective taking is not required), and a ‘privileged’ condition, in which the speaker has visual access to referents which are blocked from the view of the listener (i.e. visual perspective taking is required). In the shared condition, the use of a size adjective (e.g. ‘big cup’) is appropriate as the listener has visual access to two possible referents, whilst in the privileged condition a bare noun (e.g. ‘cup’) is appropriate as visual access to the competing cup is blocked from the listener. Table 5 shows studies utilising this

**Table 5.** Production – referential communication tasks.

Author, year	Participants			Matching			Task	Measure of reference	RC impairment in ASD group	
	Sample size (n per group)	Mean age (range) per group	Language	ASD diagnosis	Formal language	Age				Non-verbal IQ
Nadig et al. (2009)	n = 17	ASD: 11;3 (9;6–14;6) TD: (8–14)	English	SCQ/ADOS	✓ CELF d = 0.24	✓ d = 0.36	PIQ WASI d = 0.26	Director task with shared/privileged condition Mann–Whitney U No IRR	Use of size adjective in shared/privileged condition Mann–Whitney U No IRR	✓
Fukumura (2016)	n = 20	ASD-Y: 8;8 (6;8–10;9) ASD-O: 13;7 (11;8–16;3) TD-Y: 8;3 (6;3–10;5) TD-O: 13;7 (12;1–16;4) TD-Adult: 21 (18–23)	English	SCQ	✓ WASI Vocabulary d = 0.26	✓ d = 0.37	WASI Matrix reasoning d = 0.23	Director task with pictures	Use of size adjective in shared/privileged condition Logit mixed effects modelling No IRR	'Shared' condition: × 'Privileged' condition: ✓ Younger: ✓ Older: × (marginal p = .07)
Volden et al. (1997)	n = 10	ASD: 18;8 (13;6–24;4) TD: 18;8 (13;1–24;4)	English	Met DSM-III-R diagnostic criteria for autism	Reported as 'similar'	✓ d = 0	Not reported	'Guess Who' type task: provide description to identify one of two circles. VPT not required	Contrastive (i.e. correct), redundant (distinguishing feature + at least one other), uninformative (did not include distinguishing attribute) t-test, Fisher's Exact % reliability	✓
Dahlgren and Dahlgren Sandberg (2008)	n = 30	ASD: 10.06 (7.58–14.6) TD: 9.55 (7–13.92)	English	Met DSM-III-R diagnostic criteria for autism	✓ WISC VIQ d = 0.35	✓ d = 0.28	FSIQ d = 0.60 ✓ WISC PIQ d = 0.27	'Guess Who' type task: Provide description to identify one of 16 faces. VPT not required	Relevant/irrelevant/redundant features mentioned Mann–Whitney/ Wilcoxon/Spearman's No IRR	✓
Nadig et al. (2015)	n = 13	ASD: 22.2 (18–29) TD: 21;2	English	ADOS/SCQ	✓ WASI VIQ d = 0.21	✓ d = 0.33	Not reported	Referential pact paradigm with tangrams	Lexical alignment with original/new partner ANOVAs Cohen's k	Initial lexical entrainment × Adaptation to new partner ✓

ADOS: Autism Diagnostic Observation Schedule; AS: Asperger syndrome; ASD: autism spectrum disorder; ASD-O: older; ASD-Y: younger; CELF: Clinical Evaluation of Language Fundamentals; DSM: *Diagnostic and Statistical Manual of Mental Disorders*; FSIQ: Full Scale IQ; ✓: well matched; PIQ: performance IQ; SCQ: Social Communication Questionnaire; TD: typically developing; VIQ: verbal IQ; WASI: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children; X: not well matched.

paradigm to examine production of referring terms meeting our inclusion criteria.

Using a director task, Nadig, Vivanti and Ozonoff (2009: Exp 1) found that children with ASD aged 9–14 years used proportionally fewer appropriate referring terms to identify objects than a well-matched TD group (both groups:  $n=17$ ). In the privileged condition, participants with ASD tended towards over-informativity, inappropriately using a specific referring term (e.g. ‘big cup’ when there was only one cup available from the listener’s visual perspective) significantly more frequently than the TD group ( $p<.01$ ). In the shared condition, the ASD group more frequently failed to use a complex referring term when two competing referents were visible, though this group difference was only of marginal significance ( $p=.08$ , effect size  $r=0.24$ ). These findings reflect the simultaneous over and under-informativity in reference use by individuals with ASD which was also the general finding from narrative and conversational studies.

Fukumura (2015) used a similar director task whereby she directly compared the ‘privileged’ and ‘shared’ perspective conditions. The dependent variable was the percentage of complex referring expressions (e.g. ‘the small door’) as opposed to unmodified nouns (e.g. ‘the door’). Thus, if individuals with ASD were taking listener informational needs into account, there should be significantly less complex referring expressions used in the ‘privileged’ condition, since this would be over-informative from the addressee’s perspective. For both 6- to 10-year-olds (Exp 1) and 11- to 16-year-olds (Exp 2), there was a group by condition interaction, indicating that the typical controls were significantly more likely to make this audience design distinction than were the individuals with ASD.

Nadig et al. (2009) and Fukumura’s (2016) studies indicate that when the speaker and listener perspectives differ, individuals with ASD have difficulty selecting referring expressions appropriate to their listener’s perspective. However, even when participants know that their listener can see the *same* visual array as themselves, Nadig et al. (2009) found diminished performance for ASD groups. This latter observation is reflected in the findings of two referential communication studies which did not manipulate listener perspective. Both used adaptations of the original reference communication paradigm developed by Glucksberg and Krauss (1967) in which the participant and a confederate play a version of the ‘Guess Who?’ game. Volden, Mulcahy and Holdgrafer (1997) asked adolescents and adults with ASD to provide information to identify a target from one of two circles which varied on one of four possible attributes (colour, shape, pattern and position of a small black dot). Whilst individuals with ASD never failed to provide the distinguishing

feature in their description, they were more likely to include redundant information that did not uniquely identify the target referent.

Using a similar paradigm, Dahlgren and Dahlgren Sandberg (2008) asked children with ASD to provide descriptions to identify a given face from a selection of 16. They measured how many of the features mentioned were ‘*relevant*’ (appropriately discriminated between pictures), ‘*irrelevant*’ (common to all pictures e.g. ‘has a mouth’) and ‘*redundant*’ (already a given). Children with ASD produced significantly fewer relevant features than did TD controls and they also included proportionally more irrelevant than relevant features than the TD group.

Whilst director tasks have been used to manipulate visual common ground knowledge, *social* common ground (namely the ability to determine the knowledge what one shares with a specific interlocutor; Moll & Kadipasaoglu, 2013) is arguably the skill used more often when selecting referring terms in everyday conversation. In a ‘referential pact’ paradigm, Nadig, Seth and Sasson (2015) examined whether adults with ASD engaged in lexical entrainment – the process by which interlocutors come to agree on mutual referring terms. Participants provided information to enable their listener to identify one of an array of abstract forms (tangrams). Individuals tended to alter referential descriptions in co-operation with the listener over successive trials (e.g. pairs may agree to call a shape ‘the elephant’ after initially describing it as ‘a four legged or two legged animal facing the right... The head is a parallelogram and its back leg is a rectangle and the front legs look like paws’). To investigate whether this alignment of referring terms was due merely to priming or if social common ground was utilised, the game continued with either the original or a new listener. If common ground was considered, the agreed referring terms should be used with the original but not a new listener. In the ‘new listener’ condition, the ASD group were marginally ( $p=.05$ ,  $r=.37$ ) less likely than the TD group to change the referring expression (referential pact) they had agreed with the original listener.

*Studies comparing ASD with other neurodevelopmental disorders.* Some studies returned in our search compared groups of individuals with ASD to groups of children with other neurodevelopmental disorders. For five studies the disorder concerned was DLD. Such comparisons between ASD and DLD can help elucidate the degree to which formal language (lexical and morpho-syntactic skills) might be a contributing factor in proficiency with verbal reference.

One such study is that of Manolitsi and Botting (2011) (Table 6), who found that children with ASD were significantly poorer in pronominal reference use

**Table 6.** Production – studies comparing ASD with another clinical group (where no typically developing group was included in the study).

Author, year	Sample size (n per group)	Mean age (range) per group	Language	ASD diagnosis	Structural language	Age	Non-verbal IQ	Task	Measure of reference	RC impairment in ASD group
Loveland et al. (1989)	n = 13 (ASD) n = 14 (DS)	ASD: 16;2 (8–27) DS: 15;2 (12;3–18;7)	English	Not reported	✓ <u>McCarthy's Scale/PPVT</u> d = 0.26	× d = 0.20	× <u>LIPS</u> d = 1.07 (ASD higher)	Explain rules of 'zoo game' to partner	Statistical analysis and inter-rater reliability Amount of appropriate information produced and level of prompting required to elicit this t-test % agreement	✓
Loveland et al. (1990)	n = 16	ASD: 13;6 (5–27) DS: 13;3 (5–27)	English	Vineland adaptive behaviour scales	✓ <u>McCarthy's Scale/PPVT</u> d = 0.19	✓ d = 0.05	× <u>LIPS</u> d = 0.81 (ASD higher)	Retell narrative (presented as video or pup- pets) to 'naive' addressee	Character and event reference, including ambiguity, Chi-square % agreement	×
Manolitsi and Botting (2011)	n = 13	ASD: 7;2 (4;2–13) SLI: 7;4 (5–13)	Greek	Diagnosed by child psychiatrist following American Psychiatric Association (2000) criteria	× <u>CELF-Receptive</u> d = 0.97 ✓ <u>CELF-Expressive</u> d = 0.56	✓ d = 0.18	✓ d = 0.44	Story: <i>Peter and the cat</i> (Leitao and Allan, 2003)	Character reference ANOVA Cohen's k	✓

ASD: autism spectrum disorder; CELF: Clinical Evaluation of Language Fundamentals; DS: Down syndrome; LIPS: Leiter International Performance Scale; ✓: well matched; PIQ: Performance IQ; PPVT: Peabody Picture Vocabulary Test; SLI: specific language impairment; TD: typically developing; X: not well matched.

than were children with DLD during a narrative retell task ('Peter and the Cat, Leitão & Allan, 2003). Both groups were Greek-speaking children matched on expressive language, chronological age, gender and non-verbal IQ. Participants were categorised on a scale of 0–3. A child categorised as 0 would generally show great difficulty with appropriate selection of referring expressions whereas a child categorised as 3 would be a proficient user of verbal reference. There was a significant difference between the ASD and DLD group in this regard, with children with DLD showing better performance. Unfortunately, the ASD group also had lower receptive language than did the group with DLD, which makes any between-group differences in measures of verbal reference difficult to interpret.

The remaining four DLD comparison studies have already been mentioned above, since they also included a typical control group (Baltaxe & D'Angiola, 1996; Colozzo et al., 2015; Norbury & Bishop, 2003; Norbury et al., 2014). All four found significant difficulties (or effect size indicative of a difference, Norbury et al., 2014), whereby the ASD group performed worse than the DLD group. This is particularly striking in the case of the two DLD comparison studies in which the group with ASD had higher formal language scores than the group with DLD (Colozzo et al., 2015; Norbury et al., 2014). If the ASD group still showed significantly greater difficulties in verbal reference, this provides somewhat stronger evidence that formal language is unlikely to be the main cause of these pragmatic language difficulties.

In sum, the results of all five studies which compared ASD to DLD suggest that, although referential accuracy poses a challenge for children in both groups, deficits in referential communication are more pronounced in individuals with ASD, even when the latter have superior lexio-syntactic abilities. Thus, referential communication deficits in ASD are unlikely to be solely attributable to difficulties with formal language.

In addition to cross-syndrome comparisons with DLD, two studies returned in our search compared the use of reference in ASD and Down syndrome (DS), which is a neurodevelopmental disorder associated with intellectual disability (see Table 7). The first study is that of Loveland, McEvoy, Tunali and Kelley (1990) who tested children with ASD and children with DS ( $n=16$  in each group) matched on verbal mental age. Children were asked to retell a story depicted via a video or puppet show to a naïve listener. In each group an equal proportion of children made ambiguous references to characters. Whilst Loveland et al. (1990) aimed for participants to have 'similar' non-verbal IQ and chronological age, the ASD group

had marginally higher mean IQ scores and chronological age than the DS group.

In a less structured task, again comparing reference use in ASD and DS groups, Loveland, Tunalia, Mcevoy and Kelley (1989) asked participants to provide information to a naïve listener (E2) about how to play a board game. Participants were helped by E1 to provide adequate information using a gradient of prompts from more general, for example, 'Tell me about these things here' to more specific, for example, 'Tell me where to start the game'. The ASD group produced significantly less 'adequate' descriptions than did the DS group at the most 'general' level of prompting and they also required a higher level of specific prompts than did the DS group to provide the adequate amount of information. This was the case even though the two groups were matched on verbal age and although the ASD group tended towards higher non-verbal IQ scores than the DS group. Given that the ASD group had overall higher IQ in both studies, yet exhibited difficulties equal to, or more pronounced than, a DS group, these studies suggest that the ability to develop appropriate usage of verbal reference may not be due solely to latent non-verbal intellectual difficulties.

**Production summary.** Overall, our search returned 22 studies of verbal reference production in ASD. There were seven studies which compared a group with ASD to a group with another neurodevelopmental disorder (either DLD or DS) and all but one of these studies found indications of poorer performance by the ASD group.

Eleven studies compared a group with ASD to a group of typical controls, whereby groups were either not well matched for formal language and/or non-verbal IQ ability, or this was not reported (Baltaxe & D'Angiola, 1996; Colozzo et al., 2015; Dahlgren & Dahlgren Sandberg, 2008; Mäkinen et al., 2014; Nadig et al., 2015; Norbury & Bishop, 2003; Norbury et al., 2014; Novogrodsky, 2013; Novogrodsky & Edelson, 2016; Tager-Flusberg, 1995; Volden et al., 1997). Nonetheless, it is noteworthy that 10 out of these 11 studies reported that individuals with ASD performed significantly worse than typical controls on at least one reference measure.

Finally, eight studies did compare a group with ASD to well-matched controls and *all except one* (Kuijper et al., 2015) of these well-matched case-control studies found evidence of a deficit in comparison to the typical group in terms of appropriacy of verbal reference usage (Arnold et al., 2009; Banney et al., 2015; Colle et al., 2008; de Marchena & Eigsti, 2016; Fukumura, 2016; Nadig et al., 2009; Suh et al., 2014). These latter seven studies include a range of age groups. They also

**Table 7.** Comprehension – referential communication task.

Author, year	Participants		Matching			Task	Measure of reference in ASD group	RC impairment in ASD group		
	Sample size (n per group)	Mean age (range) per group	Language	ASD diagnosis	Formal language				Age	Non-verbal IQ
Volden et al. (1997)	n = 10	ASD: 18;8 (13;6–24;4) TD: 18;8 (13;1–24;4)	English	Met DSM-III-R diagnostic criteria for autism	Reported as 'similar'	✓ d = 0	Not reported	Listen as description given of circle referent	Ability to judge when sufficient information given to identify referent ×	Statistical analysis/ Inter-rater reliability
Begeer et al. (2010)	n = 34	ASD: 16;7 TD: 16;8	English	Met DSM-IV-TR diagnostic criteria for autism	WISC/WAIS Vocabulary d = 0.14	✓ d = 0.02	WISC/WAIS Arithmetic/Block Design/Picture Arrangement d = 0.14	'Real life' version of the Director Task, VPT manipulated	Egocentric errors on ambiguous trials Response latency ANCOVA Perfect IRR	Egocentric errors on ambiguous trials Response latency ANCOVA Perfect IRR
Santiesteban et al. (2015)	n = 20 (ASD) n = 18 (TD)	ASD: 36 TD: 41	English	ADOS	WAIS d = 0.15	✓ d = 0.39	WAIS d = 0.15	Computerised version of the Director Task with person/camera control. VPT manipulated	Egocentric bias in object selection Number of 100 ms fixations on the competitor object (presented in supplementary materials) ANOVA No IRR	Egocentric bias in object selection Number of 100 ms fixations on the competitor object (presented in supplementary materials) ANOVA No IRR

ADOS: Autism Diagnostic Observation Schedule; ASD: autism spectrum disorder; DSM-III: *Diagnostic and Statistical Manual of Mental Disorders, 3rd ed*; DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders, 4th ed*; ✓: well matched; TD: typically developing; VPT: Visual perspective taking; WAIS: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children; X: not well matched.



include a range of elicitation methods, namely narrative generation (Banney et al., 2015; Colle et al., 2008; Suh et al., 2014), narrative retell (Arnold et al., 2009), the ‘director’ task (Fukumura, 2016; Nadig et al., 2009) and interlocutor-specific perspective taking (de Marchena & Eigsti, 2016). Therefore, it is safe to conclude that there is very good evidence for a clear impairment in appropriate reference selection (production) in ASD.

### Comprehension of referring expressions

In contrast to the ample number of studies examining the production of referring terms in ASD, we found only three studies that compared ASD to another group in terms of comprehension of the same phenomena (see Table 7). All three suggest that the pattern of ability differs considerably between the production and comprehension of referring terms. Although not examining interpretation of referring expressions per se, Volden et al. (1997) examined the ability of the adolescents and adults with ASD to judge whether the addressee in a referential communication paradigm had sufficient information to be able to correctly identify the referent. The authors argue that their ASD group performed at ceiling on this meta-pragmatic judgement task. In fact the ASD group were correct on average 87% of the time but the typical group were correct 100% of time (with an SD of zero), with the result that statistical analyses were not carried out. Moreover, since the Glucksberg and Krauss (1967) paradigm was used (where the participant and the confederate are aware that they are viewing identical sets of cards), it could be argued that the ability to take another’s perspective was not necessary for this task since the participant merely has to judge whether a confederate’s instruction is informative from his or her *own* perspective.

The final two studies did in fact investigate performance in reference interpretation where the participant’s perspective differed from that of the speaker. Both used the director task. In contrast to the other ‘director’ studies already outlined, here the participants were in the role of the addressee. In key (ambiguous) condition trials, each participant is instructed to pick up an object (e.g. spoon) for which the participant (but crucially, not the ‘director’) can see a referential alternative (e.g. another spoon). One dependent variable is thus the number of egocentric errors made, i.e. the number of trials on which a participant selects the object which is occluded from the director’s view and thus cannot be the intended referent. A second dependent variable is typically response latency. That is, the longer a participant takes to select the correct object is an indication of

the degree to which he or she (egocentrically) considered the referential alternative as a possible target.

In the first such study, Begeer, Malle, Nieuwland and Keysar (2010) examined the ability of adolescents with ASD to interpret referring expressions (e.g. ‘the cup’ versus ‘the big cup’) when responding to instructions in a shared and a privileged condition. Across both groups, participants made egocentric errors on 39% of trials in the key (ambiguous) condition and their response latencies were also significantly longer in the ambiguous than in the neutral condition, indicating that they considered the referential alternative prior to making correct selections. However, crucially, there were no between-group differences for either of these dependent variables. This indicates that the ASD group were as able as the TD group to use visual perspective taking to interpret verbal reference, at least when the visual perspective is as simple as determining whether the interlocutor can see a particular object. Santiesteban, Shah, White, Bird and Heyes (2015) carried out a computerised version of the director task with adults. Similarly to Begeer et al. (2010), they found no between-group differences. Moreover, they found that adults with and without ASD were equally successful in completing the task when a human addressee (avatar) was replaced with a camera.

Thus, the three reference interpretation studies with typical control groups align in suggesting that the ability to take another’s perspective to accurately interpret reference is relatively spared in ASD.

### Potential cognitive underpinnings

The picture emerging from studies on the interpretation of verbal reference is that this is not an area of impairment in individuals with ASD (Begeer et al., 2010; Santiesteban et al., 2015). This stands in stark contrast to the overwhelming finding that individuals with ASD are impaired relative to both typical peers and peers with neurodevelopmental when reference *production* is examined. One possible reason for the apparent discrepancy between an impairment in the selection of an appropriate referring expression (production) and an intact ability to take another’s perspective to interpret a referring term might be the differing cognitive underpinnings of each skill. We now therefore survey studies which explicitly examined relations between referential communication in ASD and the potential cognitive underpinnings of this skill.

We begin by examining studies which have examined relationships between proficiency with verbal reference, on the one hand, and either formal language (lexical or syntactic) proficiency and/or non-verbal IQ, on the other hand, in the samples of individuals with ASD.

### *Non-verbal IQ*

Three studies (all discussed above) examined the relationship between non-verbal IQ and the appropriacy of reference selection. Nadig et al. (2009) found that performance in the shared perspective condition correlated with non-verbal IQ, that is those with higher non-verbal IQ used more adjectives when they (and their interlocutor) could see two referential alternatives (e.g. two ducks) than when only one potential referent was present. However, in the privileged perspective condition there was no relationship with non-verbal IQ, which makes the first finding difficult to interpret. In line with this latter finding, both Dahlgren and Dahlgren Sandberg (2008) and Fukumura (2016) did not find any evidence for a relationship between non-verbal IQ and any measures of reference production in their ASD groups. Thus, on the whole, these findings – when considered together with the studies outlined above comparing children with ASD to children with DS – indicate that non-verbal IQ is unlikely to play a primary causal role in difficulties with verbal reference in ASD (although analyses in future studies should certainly control for non-verbal IQ).

### *Formal language*

Three studies (all discussed above) examined the relationship between formal language and the appropriacy of reference selection. All three studies used a referential communication paradigm. Dahlgren and Dahlgren Sandberg (2008) found that verbal IQ correlated in the ASD group (but not in the TD group) with the number of relevant features mentioned and their measure of referential efficiency. However, since they did not manipulate the distinction between the participant's and the interlocutor's perspectives, it is unclear whether this indicates that formal language is important for the appropriacy of reference selection or whether it merely suggests that a more advanced mastery of formal language leads to a greater complexity of referring expressions.

The latter interpretation is supported by Fukumura (2016), who found no relationship between formal language and performance in the privileged ground condition in her ASD groups. Rather, the only relationships with formal language (British Picture Vocabulary Scale and WASI vocabulary) were with the number of adjectives produced by the ASD group in the shared ground condition. That is, those children with ASD who had larger vocabularies tended to produce more adjectives in the shared ground condition. Since the shared ground condition does not differentiate the participant's own perspective from that of the interlocutor, this finding merely indicates that those individuals

with ASD who have larger vocabularies tend to find it easier to produce complex referring expression. In contrast, Nadig et al. (2009) found that formal language ability (Clinical Evaluation of Language Fundamentals (CELF)) correlated with appropriately informative verbal reference by participants with ASD in the 'privileged view' condition of their director task, i.e. the condition which required participants to take the addressee's perspective, since it differed from their own. That is, those with higher scores on the CELF were more able to curtail the usage of complex referring expressions when this would be over-informative. In sum, it seems likely that formal language contributes to difficulties with the production of appropriate referring expressions in children with ASD. However, given that comparisons with DLD indicated that difficulties in verbal reference production are more marked in ASD, despite better formal language skills in the latter group, it appears likely that other factors may contribute to the observed impairment.

### *Theory of Mind (ToM) and executive functioning*

Traditionally, difficulties with appropriate verbal reference selection have been linked to difficulties with ToM, which is the ability to represent others' mental states including their beliefs, emotions and desires (e.g. Baron-Cohen, Leslie, & Frith, 1985). However, it is equally plausible that a failure to provide an appropriate level of information (i.e. under- and over-informativity) could be due to a failure to differentiate between old and new information during a verbal interaction (e.g. Baltaxe, 1977). Such difficulties may be caused in part by an impairment, for example, in working memory. Working memory is usually considered one component of executive functioning, which comprises a set of highly correlated, but separable, aspects of memory, inhibitory control and cognitive flexibility needed for considering consequences to actions (e.g. Miyake et al., 2000; see also Pennington & Ozonoff, 1996 for an overview of EF domains and measurement methods).

Given that several reviews and meta-analyses report clear evidence for impairments in all domains of executive functioning bar inhibitory control in ASD (e.g. Hill, 2004; Lai et al., 2016; Russo et al., 2007) and there is evidence of a link between EF and verbal reference in TD populations (e.g. Brown-Schmidt, 2009), it is somewhat surprising that only two studies examined relationships between EF and usage or interpretation of verbal reference in ASD. Both studies are also the only two to examine the relationship between the appropriacy of reference selection and ToM in ASD. The first study, carried out by Dahlgren and Dahlgren

Sandberg (2008), included only two tasks which might plausibly be considered a measure of executive functioning and both of these measured short memory. The first was ‘verbal free recall’ (11 lists of words with 10 words in each) and the second was ‘object free recall’ (in which the child is shown 10 sets of 10 objects and is required to verbally recall them). In both memory tasks, once all items had been presented, the child was asked to repeat as many words (or objects, respectively) as he or she could remember and in any order. Relationships were found between both memory measures and certain aspects of verbal reference, namely the number of relevant features mentioned and the ‘efficiency’ of reference usage, that is the extent to which descriptions were optimally informative (for the comparison group this was only significant for verbal free recall). The authors interpret this as indicating that working memory impacted on the number of referential alternatives which a child could hold in mind and possibly also on the ability to verbally encode the relevant distinguishing information

Dahlgren and Dahlgren Sandberg (2008) also directly examined the relationship between ToM and the usage/interpretation of verbal reference. To measure ToM they used a first-order change-of-location task (Baron-Cohen et al., 1985) as well as Baron-Cohen’s (1989)’s second-order false belief ‘ice-cream’ task. For the second-order ToM measure no significant relationships were found with any verbal reference measures for either group. For the children with ASD there was a relationship between first-order ToM and the same aspects of verbal reference used (number of relevant features mentioned and the ‘efficiency’) that correlated with free recall. They note, however, that the correlation with first-order ToM is based only on five children in the ASD group who failed the first-order ToM task (whereas 25 children with ASD passed). More problematically, Spearman’s rho was used for all correlational measures, when a point-biserial correlation is appropriate for the first-order ToM task which was essentially a pass/fail measure.

The other study which examined relationships in ASD between appropriacy of reference selection, on the one hand, and either EF or ToM, on the other hand, is Kuijper et al. (2015). They used the Stop Signal Reaction Time Task (Van den Wildenberg & Christoffels, 2010) to measure inhibitory control and the n-back task to measure working memory. First- and second-order ToM was assessed in a scale consisting of eight stories (Hollebrandse, Van Hout, & Hendriks, 2014). For the ‘reintroduction of character in a narrative’ condition (where a noun and not a pronoun would be appropriate), the authors found in a multivariate model relationships between reference usage and both second-order ToM and working

memory. Unfortunately, language measures were not entered into the analysis and the results were conflated over three groups, which included a group with Attention Deficit Hyperactivity Disorder, making this finding difficult to interpret.

Based on the studies included in this review, it appears there is insufficient evidence to determine whether the development of verbal reference usage in ASD is underpinned by ToM and/or EF. The degree to which ToM and EF underpin the development of verbal reference in ASD is complicated by the fact that these two areas tend to be inter-correlated with each other (e.g. Pellicano, 2013) and also with formal language (e.g. Milligan, Astington, & Dack, 2007, for a review). However, considering the evidence for relationships between both EF and ToM and other areas of pragmatics (see, e.g. Matthews, Biney, & Abbot-Smith, in press) further exploration of the cognitive underpinnings related to comprehension and production of referring terms is clearly a priority.

## Summary and discussion

The current systematic review found 19 studies which met our criteria and in which verbal reference production by a group of individuals of ASD was compared to that used by a typically developing control group. Seventeen of these 19 studies found that the group with ASD were impaired in at least one measure in terms of the appropriacy of match between context and the form of verbal reference. While many of these studies had various methodological issues, this pattern of results also held for seven of the eight studies in which the typical group were matched to the ASD group in terms of chronological age, non-verbal IQ and formal language (Arnold et al., 2009; Banney et al., 2015; Colle et al., 2008; de Marchena & Eigsti, 2016; Fukumura, 2016; Nadig et al., 2009; Suh et al., 2014). This stands in stark contrast to the findings from the three studies of verbal reference comprehension, in which individuals with ASD were observed to show typical understanding/interpretation of verbal reference. This was even true for the two studies in which the perspective of participants differed from that of the speaker and, thus, required a shift in mental perspective (Begeer et al., 2010; Sanstieban et al., 2015).

However, this apparent discrepancy between production and comprehension measures may be an artefact of certain characteristics of the existing studies rather than an actuality. The first characteristic of the data that prevents us from drawing firm conclusions is that the participants of these three comprehension studies were all adults or older adolescents; there are some indications that proficiency with the production of verbal reference may ameliorate to some degree

during adolescence (e.g. Arnold et al., 2009; de Marchena & Eigsti, 2016). However, improvement over development seems unlikely to be the full story for the difference between production and comprehension studies since two production studies with adults with ASD did find evidence of impairment in comparison to well-matched controls (Colle et al., 2008; Nadig et al., 2015).

Another possibility is that the apparent discrepancy between comprehension and production is due to task-related differences across studies. For example, the majority of production studies used narrative elicitation (for which there is no obvious comprehension-task counterpart), whereas all three comprehension studies used a referential communication paradigm (Begeer et al., 2010; Sanstieban et al., 2015; Volden et al., 1997). Indeed, one commonality amongst comprehension tasks used in all three studies (Begeer et al., 2010; Sanstieban et al., 2015; Volden et al., 1997) is that the dependent variable is binary forced choice, which is certainly far from the case for most production-dependent variables. That said, five production studies (two of which were methodologically well controlled) also used a referential communication paradigm, where the dependent variable could possibly be considered binary forced choice, and all found impairments in the ASD group relative to the typical control group (see Table 5). Therefore, it seems unlikely that the dichotomy found between comprehension and production studies can be attributed to the fact that the comprehension paradigms are binary forced choice.

To unpick the cognitive underpinnings of this discrepancy, this field needs much more fine-grained task analysis of the processes involved in the appropriate selection of referring expressions and of the processes involved in using the interlocutor's perspective to interpret referring expressions. It is tempting to suggest that production of verbal reference is inherently more burdensome to executive functioning than is interpretation of verbal reference. In production of verbal reference, the speaker requires, for example, working memory to hold information relevant to the listener whilst a sentence is formulated and executed. If the specific syntactic form of the target referring expression (e.g. simple noun phrase versus complex noun phrase) differs between trials, this may also place additional demands on mental set shifting, that is the ability to switch back and forth between multiple trials (see, e.g. Sikora, Roelofs, Hermans, & Knoors, 2016). That said, even comprehension variants of the visual perspective referential communication task have been found to tap various aspects of EF in the typical population (e.g. Brown-Schmidt, 2009; Cane, Ferguson, & Apperly, 2016; Lin et al., 2010; Nilsen & Graham, 2009). Thus, we clearly need a more precise mechanistic model of the

fine-grained steps required for comprehension and production and how this might differ depending on the specific tasks used for each.

Whatever the explanation for the discrepancy between performance on laboratory-based measures of verbal reference interpretation and laboratory-based measures of verbal reference production, there is a further overarching issue that needs to be considered when drawing conclusions about these abilities in ASD. Even if individuals with ASD are unimpaired in interpretation using the referential communication tasks, this does not mean that they are necessarily unimpaired in interpretation of verbal reference in everyday life. This is because all referential communication tasks to date in this field have essentially manipulated only level one visual perspective, which is an individual's understanding that the content of what they see may differ from the content of what another sees in the same physical position (e.g. Salatas & Flavell, 1976). This requires the ability to follow another person's line of sight and draw conclusions about whether a person's perception of an object is occluded, yet one need not have a very deep understanding of mental states to determine this (e.g. Moll & Kadipasaoglu, 2013; see also Sanstieban et al., 2015, for a sub-mentalising account).

In everyday life, in contrast, the interpretation of reference is often dependent on 'social' perspective taking/common ground, that is an understanding of what a specific interlocutor knows or is likely to find interesting or salient. This often depends on a consideration of which particular information or experiences we have shared with which specific interlocutors. The only study meeting our criteria which investigated this is Nadig et al. (2015), who found that adults with ASD were less likely than typical adults to take discussion shared with a particular interlocutor (via a referential pact) into account when selecting a referent term. Of course, there are numerous divergent ways in which social common ground can be established with a specific interlocutor. One way is through sharing a particular collaborative experience (e.g. painting an action figure) with a certain interlocutor. To date this has only been explored to a degree in a couple of very small-scale production study pilots without control groups (Geller, 1988; Rosenthal Rollins, 2014). No studies have investigated whether individuals with autism can use social common ground to *interpret* verbal reference.

## Conclusions

To move this field forward, we need studies which manipulate the role of social perspective taking and compare this using comprehension and production variants of the task in the same sample of individuals

with ASD. We also need the field to shift away from an over-reliance on narrative paradigms. In addition to some issues with narrative paradigms outlined above, narrative is problematic here because verbal reference can be used appropriately in narrative without a real consideration of the listener's perspective, by simply tracking whether the form used for introduction or maintenance of reference is appropriate from one's own perspective (see, e.g. Arnold, 2008, for a discussion of 'narrator-oriented' use of verbal reference). Finally, we need a more detailed account of how deficits revealed in experimentally elicited production of verbal reference link to pragmatic language impairments in naturalistic dialogue. To that end, it is striking that to date there exists only one case-controlled study of reference production in conversation (Baltaxe & D'Angolia, 1996) and this study had highly problematic methodological issues. We need to empirically document in more detail the degree to which an impairment in reference usage hinders real-life verbal interaction, and to demonstrate more precisely the potential links that such an impairment has with difficulties in peer interaction and/or mental health difficulties.

#### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by a Research Scholarship to the first author, who completed this study as part of her PhD thesis.

#### References

Note: References preceded by an asterisk indicate articles included in tables 2–7 of reviewed studies.

- Adams, C., Gaile, J., Freed, J., & Lockton, E. (2010). Targeted Observation of Pragmatics in Children's Conversation (TOPICC): Adapting a research tool into a clinical assessment profile. *Speech and Language Therapy in Practice*, (Spring), 7–10. Retrieved from <http://www.psych-sci.manchester.ac.uk/scip/topicc.pdf> eScholarID: 83594.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th.). Arlington, VA: American Psychiatric Publishing.
- \*Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495–527. doi:10.1080/01690960801920099.
- Arnold, J. E., Benetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism:

- Effects of discourse status and processing constraints. *Cognition*, 110(2), 131–146. doi:10.1016/j.cognition.2008.10.016.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
- Baltaxe, C. A. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology*, 2(4), 176–180. doi:10.1093/jpepsy/2.4.176.
- \*Baltaxe, C. A., & D'Angolia, N. (1996). Referencing skills in children with autism and specific language impairment. *International Journal of Language & Communication Disorders*, 31(3), 245–258. doi:10.3109/13682829609033156.
- \*Banney, R. M., Harper-Hill, K., & Arnott, W. L. (2015). The autism diagnostic observation schedule and narrative assessment: Evidence for specific narrative impairments in autism spectrum disorders. *International Journal of Speech-Language Pathology*, 17(2), 159–171. doi:10.3109/17549507.2014.977348.
- Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, 30, 285–298. doi:10.1111/j.1469-7610.1989.tb00241.x.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1), 37–46. doi:10.1016/0010-0277(85)90022-8.
- \*Begeer, S., Malle, B. F., Nieuwland, M. S., & Keysar, B. (2010). Using theory of mind to represent and take part in social interactions: Comparing individuals with high-functioning autism and typically developing controls. *European Journal of Developmental Psychology*, 7(1), 104–122. doi:10.1080/17405620903024263.
- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bishop, D. V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2. Terminology. *PeerJ Preprints*, 4, e2484v1.
- Brandt, S. (2011). The form and function of relative clauses in discourse and experimental studies with children. In E. Kidd (Ed.), *The acquisition of relative clauses: Processing, typology and function*. John Benjamins Publishing Company.
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, 16(5), 893–900. doi:10.3758/PBR.16.5.893.
- Cane, J. E., Ferguson, H. J., & Apperly, I. (2016). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 591–610. doi:10.1037/xlm0000345.
- Chiat, S. (1986). Personal pronouns. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 339–355). Cambridge, UK: Cambridge University Press.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber &

- I. A. Sag (Eds.), *Elements of discourse understanding*. Cambridge, UK: Cambridge Univ. Press.
- Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. Leny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam, UK: North-Holland.
- \*Colle, L., Baron-Cohen, S., Wheelwright, S., & van Der Lely, H. K. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, 38(1), 28. doi:10.1007/s10803-007-0357-5.
- \*Colozzo, P., Morris, H., & Mirinda, P. (2015). Narrative production in children with autism spectrum disorder and specific language impairment. *Canadian Journal of Speech-Language Pathology & Audiology*, 39(4), 316–332.
- Constantino, J. N., & Gruber, C. P. (2007). *Social responsiveness scale (SRS)*. Los Angeles, CA: Western Psychological Services.
- \*Dahlgren, S., & Dahlgren Sandberg, A. D. (2008). Referential communication in children with autism spectrum disorder. *Autism: The International Journal of Research and Practice*, 12(4), 335–348. doi:10.1177/1362361308091648.
- \*de Marchena, A., & Eigsti, I. (2016). The art of common ground: Emergence of a complex pragmatic language skill in adolescents with autism spectrum disorders. *Journal of Child Language*, 43(1), 43–80. doi:10.1017/S0305000915000070.
- \*Fukumura, K. (2016). Development of audience design in children with and without ASD. *Developmental Psychology*, 52(1), 71–87. doi:10.1037/dev0000064.
- Geller, E. (1988). The interplay between linguistic and social-cognitive knowledge in perspective-taking by autistic children. *Communication Disorders Quarterly*, 14(1), 23–44. doi:10.1177/152574018801400103.
- Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies*, 6(3), 463–481. doi:10.1075/is.6.3.10ger.
- Girouard, P., Ricard, M., & Decarie, T. (1997). The acquisition of personal pronouns in French-speaking and English-speaking children. *Journal of Child Language*, 24(2), 311–326.
- Glucksberg, S., & Krauss, R. M. (1967). What do people say after they have learned how to talk? Studies of the development of referential communication. *Merrill-Palmer Quarterly of Behavior and Development*, 13(4), 309–316.
- Graf, E., & Davies, C. (2014). The production and comprehension of referring expressions. In D. Matthews (Ed.), *Pragmatic development in first language acquisition*. Amsterdam, UK/Philadelphia, PA: John Benjamins Publishing Company.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics. Volume 3: Speech acts* (pp. 225–242). New York, NY: Seminar Press.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2), 274–307. Retrieved from <http://www.jstor.org/stable/416535>.
- Geurts, H. M., Vert , S., Oosterlaan, J., Roeyers, H., & Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of child psychology and psychiatry*, 45(4), 836–854.
- Happ , F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36(1), 5–25. doi:10.1007/s10803-005-0039-0.
- Helland, W. A., Lundervold, A. J., Heimann, M., & Posserud, M. B. (2014). Stable associations between behavioral problems and language impairments across childhood – The importance of pragmatic language problems. *Research in Developmental Disabilities*, 35(5), 943–951. doi:10.1016/j.ridd.2014.02.016.
- Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, 8(1), 26–32. doi:10.1016/j.tics.2003.11.003.
- Hollebrandse, B., Van Hout, A., & Hendriks, P. (2014). Children’s first and second order false belief reasoning in a verbal and a low-verbal task. *Synthese*, 191(3), 321–333.
- Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form and use in context: Linguistics applications* (pp. 11–42). Washington, DC: Georgetown University Press.
- Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4(1), 71–84. doi:10.1515/IP.2007.004.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38. doi:10.1111/1467-9280.00211.
- Kidd, E., & Cameron-Faulkner, T. (2008). The acquisition of the multiple senses of with. *Linguistics*, 46(1), 33–61. doi:10.1515/LING.2008.002.
- \*Kuijper, S., Hartman, C., & Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PLoS One*, 10(7), e0132408. doi:10.1371/journal.pone.0132408.
- Lai, C. L. E., Lau, Z., Lui, S. S., Lok, E., Tam, V., Chan, Q., . . . Cheung, E. F. C. (2016). Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. *Autism Research*, 10(5), 911–939. doi:10.1002/aur.1723.
- Landa, R. (2000). Social language use in Asperger’s syndrome and high-functioning autism. In A. Klin, F. R. Volkmar & S. S. Sparrow (Eds.), *Asperger’s syndrome* (pp. 125–155). New York, NY: Guilford Press.
- Leit o, S., & Allan, L. (2003). *Peter and the cat: Narrative assessment*. Keighley, UK: Black Sheep Press.
- Levinson, S. (1989). A review of relevance. *Journal of Linguistics*, 25, 455–472.
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mind-blind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46(3), 551–556. doi:10.1016/j.jesp.2009.12.019.

- Lind, S. E., Williams, D. M., Bowler, D. M., & Peel, A. (2014). Episodic memory and episodic future thinking impairments in high-functioning autism spectrum disorder: An underlying difficulty with scene construction or self-projection? *Neuropsychology*, *28*(1), 55–67. doi:10.1037/neu0000005.
- \*Loveland, K. A., McEvoy, R. E., Tunali, B., & Kelley, M. L. (1990). Narrative story telling in Autism and Down's syndrome. *British Journal of Developmental Psychology*, *8*(1), 9–23. doi:10.1111/j.2044-835X.1990.tb00818.x.
- \*Loveland, K. A., Tunalia, B., Mcevoy, R. E., & Kelley, M. L. (1989). Referential communication and response adequacy in autism and Down's syndrome. *Applied Psycholinguistics; Applied Psycholinguistics*, *10*(3), 301–313. doi:10.1017/S014271640000864X.
- \*Mäkinen, L., Loukusa, S., Leinonen, E., Moilanen, I., Ebeling, H., & Kunnari, S. (2014). Characteristics of narrative language in autism spectrum disorder: Evidence from the Finnish. *Research in Autism Spectrum Disorders*, *8*(8), 987–996. doi:10.1016/j.rasd.2014.05.001.
- \*Manolitsi, M., & Botting, N. (2011). Language abilities in children with autism and language impairment: Using narrative as a additional source of clinical information. *Child Language Teaching and Therapy*, *27*(1), 39–55. doi:10.1177/0265659010369991.
- Matthews, D., Biney, H., & Abbot-Smith, K. Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition and executive functions. *Language Learning and Development*, (in press).
- Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Press.
- Milligan, K., Astington, J., & Dack, L. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi:10.1006/cogp.1999.0734.
- Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience*, *7*, 558. doi:10.3389/fnhum.2013.00558.
- Moore, R. (2014). Ontogenetic constraints on Paul Grice's theory of communication. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 87–104). Amsterdam, UK: John Benjamins.
- Murphy, S., Faulkner, D., & Farley, L. (2014). The behaviour of young children with social communication disorders during dyadic interaction with peers. *Journal of Abnormal Child Psychology*, *42*(2), 277–289. doi:10.1007/s10802-013-9772-6.
- \*Nadig, A., Seth, S., & Sasson, M. (2015). Global similarities and multifaceted differences in the production of partner-specific referential pacts by adults with autism spectrum disorders. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01888.
- \*Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Research*, *2*(6), 334–347. doi:10.1002/aur.102.
- Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*(2), 220–249. doi:10.1016/j.cogpsych.2008.07.002.
- Norbury, C. (2014). Practitioner review: Social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry*, *55*(3), 204–216. doi:10.1111/jcpp.12154.
- \*Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287–313. doi:10.1080/136820310000108133.
- \*Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language*, *41*(3), 485–510. doi:10.1017/S030500091300007X.
- \*Novogrodsky, R. (2013). Subject pronoun use by children with autism spectrum disorders (ASD). *Clinical Linguistics & Phonetics*, *27*(2), 85–93. doi:10.3109/02699206.2012.742567.
- \*Novogrodsky, R., & Edelson, L. R. (2016). Ambiguous pronoun use in narratives of children with autism spectrum disorders. *Child Language Teaching and Therapy*, *32*(2), 241–252. doi:10.1177/0265659015602935.
- Pellicano, E. (2013). Testing the predictive power of cognitive atypicalities in autistic children: Evidence from a 3-year follow-up study. *Autism Research*, *6*, 258–267. doi:10.1002/aur.1286.
- Pennington, B., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*(1), 51–87. doi:10.1111/j.1469-7610.1996.tb01380.x.
- Perkins, M. R. (2005). Pragmatic ability and disability as emergent phenomena. *Clinical Linguistics & Phonetics*, *19*(5), 367–377. doi:10.1080/02699200400027155.
- Renfrew, C. E. (1991). *The bus story: A test of narrative speech*. Oxford, UK: Winslow Press.
- Rosenthal Rollins, P. (2014). Narrative skills in young adults with high-functioning autism spectrum disorders. *Communication Disorders Quarterly*, *36*(1), 21–28. doi:10.1177/1525740114520962.
- Russo, N., Flanagan, T., Iarocci, G., Berringer, D., Zelazo, P., & Burack, J. (2007). Deconstructing executive deficits among persons with autism: Implications for cognitive neuroscience. *Brain and Cognition*, *65*(1), 77–86.
- Salatas, H., & Flavell, H. (1976). Perspective taking: The development of two components of knowledge. *Child Development*, *47*, 103–109. Retrieved from <http://www.jstor.org/stable/1128288>.
- \*Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control.

- Psychonomic Bulletin & Review*, 22(3), 844–849. doi:10.3758/s13423-014-0716-0.
- Sikora, K., Roelofs, A., Hermans, D., & Knoors, H. (2016). Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. *The Quarterly Journal of Experimental Psychology*, 69(9), 1719–1740. doi:10.1080/17470218.2015.1093007.
- Simmons, E. S., Paul, R., & Volkmar, F. (2014). Assessing pragmatic language in autism spectrum disorder: The Yale in vivo pragmatic protocol. *Journal of Speech, Language, and Hearing Research*, 57(6), 2162–2173. doi:10.1044/2014\_JSLHR-L-14-0040.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd.). Oxford, UK: Blackwell.
- \*Suh, J., Eigsti, I., Naigles, L., Barton, M., Kelley, E., & Fein, D. (2014). Narrative performance of optimal outcome children and adolescents with a history of an autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders*, 44(7), 1681–1694. doi:10.1007/s10803-014-2042-9.
- \*Tager-Flusberg, H. (1995). ‘Once upon a rabbit’: Stories narrated by autistic children. *British Journal of Developmental Psychology*, 13(1), 45–59. doi:10.1111/j.2044-835X.1995.tb00663.x.
- Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd., pp. 335–364). Hoboken, NJ: Wiley and Sons.
- van den Wildenberg, W. P., & Christoffels, I. K. (2010). STOP TALKING! Inhibition of speech is affected by word frequency and dysfunctional impulsivity. *Frontiers in Psychology*, 1, 145. doi:10.3389/fpsyg.2010.00145.
- \*Volden, J., Mulcahy, R. F., & Holdgrafer, G. (1997). Pragmatic language disorder and perspective taking in autistic speakers. *Applied Psycholinguistics*, 18(2), 181–198. doi:10.1017/S0142716400009966.
- Wiesner, D. (1991). *Tuesday*. New York, NY: Clarion Books.