



ELSEVIER

Contents lists available at ScienceDirect

Cognition

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

Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development

Calum Hartley^{a,*}, Laura-Ashleigh Bird^{a,1}, Padraic Monaghan^{a,b}

^a Department of Psychology, Lancaster University, Lancaster LA1 4YF, United Kingdom

^b Department of English, University of Amsterdam, Amsterdam, P.O Box 1641, Netherlands

ARTICLE INFO

Keywords:

Autism spectrum disorder
Word learning
Cross-situational learning
Retention
Generalisation
Typical development

ABSTRACT

Word learning is complicated by referential ambiguity – there are often multiple potential targets for a newly-heard word. While typically developing (TD) children can accurately infer word meanings from cross-situational statistics, specific difficulties tracking word-object co-occurrences may contribute to language impairments in autism spectrum disorder (ASD). Here, we investigate cross-situational word learning as an integrated system including mapping, retention, and generalisation in both typical development and autism. In Study 1, children with ASD were as accurate at disambiguating the meanings of novel words from statistical correspondences as TD controls matched on receptive vocabulary. In Study 2, both populations spontaneously utilised social and non-social attentional cues to facilitate and accelerate their mapping of word-referent relationships. Across Studies 1 and 2, both groups retrieved and generalised word-referent representations with impressive and comparable accuracy. Although children with ASD performed very similarly to TD children on measures of learning accuracy, they were significantly slower to identify correct referents under both cued and non-cued learning conditions. These findings indicate that mechanisms supporting cross-situational word learning, and the relationships between them, are not qualitatively atypical in language-delayed children with ASD. However, the increased time required to generate correct responses suggests that these mechanisms may be less efficient, potentially impacting learning in natural environments where visual and auditory stimuli are presented rapidly. Our data support claims that word learning in the longer term is driven by the gradual accumulation of word-object associations over multiple learning instances and could potentially inform the development of interventions designed to scaffold word learning.

1. Introduction

Word learning is of crucial importance to children's language acquisition (Patael & Diesendruck, 2008). When an unfamiliar word is segmented from speech, mapping the phonological form to its meaning involves: (a) identification of intended meaning (referent selection), (b) storage of the word-referent association enabling later retrieval (retention), and (c) appropriate extension to new members of the same semantic category (generalisation). In natural learning environments, these processes are complicated by referential ambiguity. Parents direct 300–400 words at their children per hour (Hart & Risley, 1995) and each unfamiliar word could refer to innumerable unknown referents (Cartmill et al., 2013; Quine, 1960). One way that children resolve this ambiguity is by learning across situations, aggregating data across naming events to form stable word-referent relationships (MacDonald,

Yurovsky, & Frank, 2017; McMurray, Horst, & Samuelson, 2012; Saffran, Aslin, & Newport, 1996; Smith & Yu, 2008; Vlach & DeBrock, 2019).

Unlike typically developing (TD) children, many children with autism spectrum disorder (ASD) have difficulty learning words. Children with ASD are often significantly delayed in their production of first words (Howlin, 2003), and approximately 25–30% have minimal-to-no language during childhood (Anderson et al., 2007; Norrelgen et al., 2015; Rose, Trembath, Keen, & Paynter, 2016). Here, we investigate whether cross-situational word learning is atypical in children with ASD, and how differences in this ability may affect lexical retention and generalisation.

Fast mapping, evident in early stages of word learning, occurs when a child hears a novel word and uses the linguistic and non-linguistic context to rapidly acquire information about its meaning (e.g. mapping

* Corresponding author.

E-mail addresses: c.hartley@lancaster.ac.uk (C. Hartley), l.bird@ucl.ac.uk (L.-A. Bird), p.monaghan@lancaster.ac.uk (P. Monaghan).

¹ Laura-Ashleigh Bird's present address is Institute of Cognitive Neuroscience, UCL, WC1E 6BT, United Kingdom.

<https://doi.org/10.1016/j.cognition.2020.104265>

Received 15 July 2019; Received in revised form 5 March 2020; Accepted 6 March 2020

Available online 04 April 2020

0010-0277/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

it to an unfamiliar object; Carey, 1978). However, in the face of referential uncertainty, accurate fast mapping may be impossible. Under such circumstances, correct word-referent mappings can be reliably acquired over multiple ambiguous naming events (Yurovsky & Frank, 2015). If a child hears an unfamiliar word (“cat”) while viewing two unfamiliar objects (a cat and a mouse), they have no way of identifying the correct referent without external input. But, if the child hears the word again while viewing a different congregation of objects (a cat and a dog), they could combine word-object co-occurrence frequencies across exposures and accurately infer the meaning of “cat” (Smith & Yu, 2008). Empirical evidence demonstrates that TD adults and children spontaneously utilise word-object co-occurrence frequencies to identify the meanings of unknown words (Monaghan & Mattock, 2012; Roembke & McMurray, 2016; Smith & Yu, 2008; Vlach & Johnson, 2013; Vouloumanos & Werker, 2009; Yu & Smith, 2007; Yurovsky, Fricker, Yu, & Smith, 2014). For example, Smith and Yu (2008) presented TD 12- and 14-month-olds with pairs of novel words in conjunction with pairs of unfamiliar referents. Although these trials were individually ambiguous, each word reliably co-occurred with a single referent. Immediately after training, children were presented with test trials in which a single word was presented in conjunction with the target referent and a distracter. The results revealed that infants looked significantly longer at target referents, indicating their ability to infer correct meanings for multiple words based on cross-situational statistics alone.

Although there is substantial evidence that TD children's sensitivity to statistical associations between words and objects supports the acquisition of word-referent mappings, the underlying processes have been subject to debate (MacDonald et al., 2017). McMurray et al. (2012) propose that cross-situational word learning results from dynamically strengthening and weakening inter-connections between words and multiple potential referents within a system. By contrast, Trueswell and colleagues propose that learners map words to single referents, and then verify or update their hypotheses based on new evidence (Trueswell, Medina, Hafri, & Gleitman, 2013). Alternatively, both mechanisms may be employed depending on learning and environmental constraints (e.g. number of words, targets, and competitors; Yurovsky & Frank, 2015).

When faced with ambiguous naming events, TD children also infer meaning from non-linguistic social cues including gaze, gesture, and affect (Baldwin, 1993; Nappa, Wessel, McEldoon, Gleitman, & Trueswell, 2009; Tomasello, Strosberg, & Akhtar, 1996). As words often refer to whatever referent is currently the focus of a speaker's attention (Bloom, 2002), following social cues can facilitate statistical learning by increasing the frequency of exposure to correct word-referent co-occurrences, thereby increasing children's uptake from input (McMurray et al., 2012; Monaghan, 2017). Although research has yet to directly investigate how children utilise social cues in the service of cross-situational word learning, MacDonald et al. (2017) explored this relationship in neurotypical adults. They found that when referential ambiguity was reduced by eye gaze towards targets, participants were quicker and more accurate in their learning, and reduced their tracking of alternative possible mappings between words and objects.

To date, studies of cross-situational learning have focused predominantly on how TD children and adults identify referential meaning. As such, participants' understanding of word-referent mappings is conventionally tested immediately after training. However, accurate referent selection does not constitute word learning *per se*. Horst and colleagues have repeatedly demonstrated that TD children forget new words just 5 min after performing at ceiling on fast mapping tasks (Axelsson, Churchley, & Horst, 2012; Horst & Samuelson, 2008; Horst, Scott, & Pollard, 2010). Under those learning circumstances, though children can select the novel referent accurately, their retention may be poor due to the paucity of associative information about the mapping. Thus, it is necessary to investigate how novel word-referent relationships that are acquired through cross-situational learning are

retained after a delay. As theoretical accounts of lexical retention emphasise the role of associative mechanisms in gradually accumulating knowledge over multiple exposures (e.g. McMurray et al., 2012), we may expect cross-situational mappings to yield more robust and long-lasting encoding than fast mapping as word-referent associations are accumulated across several naming events. Indeed, Vlach and DeBrock (2019) showed that TD children aged 47 months and older could retain words learned via cross-situational statistics at above-chance accuracy following a 5-minute delay.

Furthermore, studies have yet to explore how cross-situational learning supports extension of labels to unseen referents. As basic nouns refer to categories, children must learn to appropriately generalise labels to exemplars that have not been directly named. By approximately 24-months, TD children infer that noun-referent relationships are constrained by shape, and thereby generalise newly-learned words based on this feature rather than other perceptual properties (e.g. colour, size, texture; Landau, Smith, & Jones, 1988). In a recent study by Chen, Zhang, and Yu (2018), neurotypical adults were trained on associations between words and objects, but also words and a category name that extended over several related instances for an object. Participants were able to learn both individual labels and category labels. However, this study did not test whether use of a word extended to additional, unseen members of the category. Therefore, a novel goal of this study is to examine cross-situational word learning as a system that includes a dynamic mapping process, retention of meaning, and generalisation to novel referents.

Word learning deficits are a critical feature within the broader constellation of language impairments that characterise ASD (see Arunachalam & Luyster, 2016). Previous research has focussed extensively on how ASD affects children's ability to identify referents for novel words within individual naming events. On one hand, many studies report that fast mapping in language-delayed children with ASD is not reliably directed by ostensive social cues (e.g. Baron-Cohen, Baldwin, & Crowson, 1997; Gliga et al., 2012; Preissler & Carey, 2005). In contrast, numerous other studies have shown that children across the spectrum reliably utilise lexical heuristics, such as mutual exclusivity, to guide their referent selection (Bedford et al., 2013; de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011; Hartley, Bird, & Monaghan, 2019; Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, & Tager-Flusberg, 2007; Preissler & Carey, 2005).

However, understanding of how ASD affects children's ability to infer word-referent relationships in the absence of informative cues is extremely limited. Only two studies, to our knowledge, have assessed cross-situational word mapping in ASD. In McGregor, Rost, Arenas, Farris-Trimble, and Stiles (2013), children with ASD and TD controls matched on age (TD $M = 10.9$ years; ASD $M = 11.2$ years), nonverbal IQ (TD $M = 106$; ASD $M = 108$), and receptive vocabulary (TD M standard score = 113; ASD $M = 109$) mapped novel words to unfamiliar objects presented in sets of three. An actor gazed at the target object on half of the trials for each word (‘facilitative trials’) and gazed forward for the other half (‘neutral trials’). Hence, on neutral trials, children had to rely on cross-trial information to accurately map the novel words (although, for some words, referents may have already been unambiguously identified in preceding facilitative trials). Both groups achieved above-chance accuracy on neutral trials, and significantly benefited from the ostensive social cues provided in facilitative trials. In a recent eye-tracking study, Venker (2019) assessed children with ASD aged 4–7.9 years and TD children aged 2.6–7.9 years matched on receptive vocabulary (TD $M = 6.3$ years; ASD $M = 5.8$ years) on a cross-situational mapping task modelled on Smith and Yu's (2008) paradigm. During training, children were exposed to four novel words, presented in pairs in conjunction with pairs of unfamiliar objects. Children's understanding of the novel word-object mappings was tested immediately after training. Both children with ASD and TD children looked longer at correct targets than expected by chance, and no between-population differences in accuracy were

identified.

The lack of population differences in these studies aligns with recent evidence that ASD does not impair statistical learning when processing visual stimuli (Foti, De Crescenzo, Vivanti, Menghini, & Vicari, 2015; Roser, Aslin, McKenzie, Zahra, & Fiser, 2015) or segmenting speech (Haebig, Saffran, & Ellis Weismer, 2017; Mayo & Eigsti, 2012; Obeid, Brooks, Powers, Gillespie-Lynch, & Lum, 2016). However, a potentially more sensitive measure of the processing involved in cross-situational word learning is the speed at which children can access meaning (Kohnert, Bates, & Hernandez, 1999). In cross-situational word learning paradigms, words and objects tend to be presented at a regular pace during training and testing. However, in natural language learning, words and their referents tend to co-occur at a substantially faster rate with more noise and distractions. Thus, overall accuracy may not be a true reflection of children's ability to associate words with meanings given the processing bottleneck that is fast, continuous, online speech (Christiansen & Chater, 2016). Here, we examine the impact of autism on both cross-situational learning accuracy and response times, while also exploring the influence of individual differences in chronological age, receptive vocabulary, and non-verbal intelligence.

It is also important to recognise that identification of meaning is just one component of word learning (McMurray et al., 2012) and it is currently unknown how effectively children with ASD are able to retain and generalise new words that are acquired through tracking cross-situational statistics. Unlike referent selection, relatively little is known about lexical retention in ASD. Norbury, Griffiths, and Nation (2010) found that high-functioning children with ASD were unimpaired in their ability to retain word-object associations, but they remembered significantly less semantic information about referents over time (see also Henderson, Powell, Gareth Gaskell, & Norbury, 2014). Bedford et al. (2013) discovered that ostensive social feedback enhanced retention of novel word meanings for TD 2-year-olds, but not those at high-risk of developing ASD. Together, these studies suggest that consolidation of word-referent relationships may be atypical in ASD. Regarding generalisation, many children with ASD struggle to extend information and behaviours across different contexts (Happé & Frith, 2006), and this difficulty can manifest in the extension of verbal labels to novel referents. Tek, Jaffery, Fein, and Naigles (2008) and Potrzeba, Fein, and Naigles (2015) found that preschoolers with ASD did not reliably extend novel labels on the basis of shape. Hartley and Allen (2014) reported that language-impaired children with ASD frequently extended labels to novel referents on the basis of shape (a category-defining cue) or colour (a category-irrelevant cue). These findings suggest that early lexical development in ASD could be hindered by the overgeneralisation of labels based on category-irrelevant cues.

Only a single study to date has explored how children with ASD identify, retain, and generalise new words within the same experimental task. In, Hartley et al. (2019), language-delayed children with ASD and TD children (ASD M age = ~8 years; TD M age = ~5 years) matched on receptive vocabulary (TD M = ~5 years; ASD M = ~5 years) identified the names of novel objects via mutual exclusivity in a standard fast mapping task, before completing tests of retention and generalisation after a 5-min delay. Participants in study 1 received no feedback while participants in study 2 received either social feedback (the experimenter looked and pointed towards the target object, while repeating its label) or non-social feedback (the experimenter activated a flashing light underneath the target object and repeated its label while looking towards the participant) after selecting referents in the fast mapping task. Across studies, both children with ASD (84–97%) and TD children (98–100%) fast mapped novel word-referent pairings at above-chance (33%) accuracy (although the between-group difference was statistically significant), but were relatively less accurate on measures of delayed retention (ASD: 41–57%; TD: 42–44%) and generalisation (ASD: 41–59%; TD: 34–46%). Surprisingly, children with ASD who received social feedback during fast mapping achieved the most accurate retention and generalisation, outperforming

TD controls in the same condition and children with ASD who received non-social feedback or no feedback. The disparity between referent selection and delayed test performance in both populations emphasises the importance of studying word learning as a system of integrated processes that occur over short and long timescales. Based on these results, Hartley et al. (2019) propose that fundamental word learning mechanisms, and the relationships between them, are not qualitatively impaired by ASD (when expectations are based on language development rather than chronological age).

If Hartley et al.'s (2019) proposal is correct, then we may anticipate that children with ASD would perform as accurately as vocabulary-matched TD controls on measures of cross-situational word learning and subsequent retention/generalisation. Moreover, we may expect that children with ASD could benefit from the provision of social cues that reduce referential ambiguity. However, potential deficits in retention may be obscured in fast mapping studies because the likelihood of TD children forgetting labels is relatively high. Between-population differences may be more likely to emerge following cross-situational learning where the increased provision of word-object associations could provide a more robust scaffold for retention in typical development. Hartley et al. (2019) hypothesise that children with ASD benefitted from social feedback because it was provided after they had mapped word-referent relationships. Having already figured out what the novel words represented, children did not need to infer communicative intent from the experimenter's cues. By contrast, ASD may impact cross-situational word learning by reducing the likelihood that children utilise available social cues to overcome referential ambiguity, diminishing the accuracy and speed of encoding correct word-referent relationships. Indeed, it is possible that children with ASD may follow social cues but fail to use them as a basis for establishing referential meaning (Gluga et al., 2012). Should this be the case, it might be possible to increase uptake of cross-situational linguistic input by providing highly-salient *non-social* attentional cues that direct children's attention to target referents during training.

This study is the first to systematically investigate whether children with ASD and concomitant language delays are capable of mapping, retaining, and generalising new words based on cross-situational statistics. By examining cross-situational word learning as a multi-stage process we identify how mechanisms are related in typical development and highlight the nature, and location(s), of potential weaknesses in ASD. To assess how individual differences influence these processes, we recruited children with ASD with varying degrees of language delay (relative to their chronological age). Our samples had receptive vocabulary age equivalents around 5–6 years on average. We targeted a similar age group as Suanda, Mugwanya, and Namy's (2014) study of cross-situational word learning in typical development, as this enables us to understand the processes involved at the point where children's vocabulary is beginning to rapidly increase. Furthermore, we know that children with ASD with these language skills are able to identify and retain words through fast mapping (Hartley et al., 2019), but we do not know if they are successful when learning under ambiguous circumstances via cross-situational statistics.

In Study 1, children with ASD and TD controls matched on receptive vocabulary were presented with pairs of unfamiliar objects (via a touchscreen tablet) and instructed to identify the referent of a novel word. These exposures were intentionally ambiguous (i.e., there was no cue to which object was the correct referent), but children could disambiguate word-object pairings based on cross-trial statistics. After a 5-minute delay, children's retention and generalisation of the novel words were assessed. Based on previous cross-situational learning studies (e.g. Venker, 2019), we predicted that children with ASD would map novel word-referent pairings as accurately as vocabulary-matched controls during training. Based on Hartley et al. (2019), we tentatively predicted that the two populations would respond similarly on delayed tests of retention and generalisation. As theoretical accounts of retention emphasise the importance of accumulating word-object correspondences

over multiple exposures (e.g. McMurray et al., 2012), we anticipated that cross-situational learning would elicit more accurate retention and generalisation than fast mapping. We also measured participants' response times as there is evidence that reaction times for linguistic stimuli may be slowed for children with ASD in comparison to TD children, despite similar mapping accuracy (e.g. Bavin, Kidd, Prendergast, & Baker, 2016). Consequently, we predicted that children with ASD may be slower to identify correct referents than TD controls.

In Study 2 we tested whether cross-situational word learning improves when referential ambiguity is potentially diminished by social and non-social attentional cues. Importantly, this research will advance theoretical understanding of language acquisition by showing how cross-situational word mapping relates to retention and generalisation in both typical development and ASD.

2. Study 1: cross-situational word mapping, retention, and generalisation in typical development and autism

2.1. Method

2.1.1. Participants

Participants were 15 children with ASD (13 males, 2 females; M age = 8.78 years; SD = 2.92 years) recruited from specialist schools, and 16 TD children (7 males, 9 females; M age = 5.52 years; SD = 1.12 years) recruited from mainstream schools and nurseries (see Table 1). Groups were matched on receptive vocabulary as measured by the British Picture Vocabulary Scale (BPVS; ASD: M age equivalent = 5.35 years, SD = 2.17; TD: M age equivalent = 5.84 years, SD = 1.20; Dunn, Dunn, Whetton, & Burley, 1997), $t(29) = 0.78$, $p = .44$. Receptive vocabulary was selected as the primary matching criterion because it reflects children's ability to learn word-referent relationships – the linguistic ability at the core of this research (Bion, Borovsky, & Fernald, 2013; Kalashnikova, Mattock, & Monaghan, 2016). All children had normal or corrected-to-normal colour vision. Children with ASD were previously diagnosed by a qualified educational or clinical psychologist, using standardised instruments (i.e. Autism Diagnostic Observation Scale and Autism Diagnostic Interview – Revised; Lord, Rutter, & Le Couteur, 1994; Lord, Rutter, DiLavore, & Risi, 2002) and expert judgement. Diagnoses were confirmed via the Childhood Autism Rating Scale 2 (CARS; Schopler, Van Bourgondien, Wellman, & Love, 2010), which was completed by each participant's class teacher (ASD M score = 36.43, SD = 6.42; TD M score = 15.28, SD = 0.41). Children with ASD were significantly older ($t(29) = 4.16$, $p < .001$, $d = 1.48$), and had significantly higher CARS scores ($t(29) = 13.09$, $p < .001$, $d = 4.65$) than the TD children. Children's non-verbal intellectual abilities were measured using the Leiter-3 (Roid, Miller, Pomplun, & Koch, 2013). The mean IQ for the ASD group was 83.73 (SD = 19.54), and the mean IQ of the TD group was significantly

higher at 102.38 (SD = 6.16), $t(29) = 3.63$, $p = .001$, $d = 1.27$. However, the groups' raw scores on the Leiter-3 did not significantly differ (ASD M score = 70.93, SD = 18.89; TD M score = 69.69, SD = 13.51), $t(29) = 0.21$, $p = .83$, indicating that their non-verbal cognitive abilities were similar at time of testing (when age was not considered). In comparison to most prior word learning research with TD children, our samples were older and more advanced in terms of receptive language development. Thus, we can be confident that any between-population differences in word learning are not the consequence of insufficient language skills in the ASD samples. All procedures performed in this study (Study 1 and Study 2) involving human participants were in accordance with the ethical standards of institutional and national research committees. Informed consent was obtained from parents/caregivers prior to children's participation.

Fifteen TD children and 13 children with ASD who participated in this study also participated in Hartley et al. (2019). The experimental tasks for that study and this study were administered one week apart (in a random order), so participants were not re-tested on the battery of standardised assessments. The two studies differed in terms of the novel words and objects used, how the experimental tasks were delivered (presenting real objects vs. a tablet computer), and the learning mechanisms being studied (fast mapping vs. cross-situational learning). Given these differences, and the time delay between administration of the two studies, we were confident that children's performance would not be influenced by interference effects.

2.1.2. Materials

Stimuli included six novel words selected from the NOUN database (Horst & Hout, 2016) and colour photographs of twelve unfamiliar objects. The novel words (virdex, nelby, gasser, blicket, fiffin, teebu) were recorded by a female speaker. The pictures of unfamiliar objects were divided into six sets. Each set included a 'named object' (which was paired with a novel word during training) and a 'shape match' (a differently-coloured variant of the named object that was presented in generalisation trials). Each shape match was the same colour as a named object from a different set, and therefore served as a 'colour match' in generalisation trials. The relationship between shape and colour of object stimuli within and across sets is illustrated in Fig. 1. Colour photographs of 12 familiar objects (cat, shoe, bed, tree, cow, banana, duck, dog, apple, flower, car, ball) were used in warm-up trials. Familiar objects were selected on the basis that most children understand their linguistic labels by 15-months (Fenson, Marchman, Thal, Reznick, & Bates, 2006). All stimuli were presented via a Microsoft Surface Pro 4 tablet computer. Pictures of objects were approximately 4 cm × 4 cm in size and separated on screen by 10 cm. As similar stimuli were used in Hartley et al. (2019), we were confident that our materials had the potential to be learnable, and generalisable, by the participants in this study.

Table 1

Characteristics of samples in Study 1 and Study 2 (standard deviation in parentheses).

Population	Condition	N	Gender	Chron. age (years)	BPVS age equiv. (years)	CARS score	Leiter-3 raw score
TD	No cue (Study 1)	16	7 males 9 females	5.52 (1.12; 3.92–8.17)	5.84 (1.20; 4.17–12.00)	15.38 (0.50; 15.00–16.00)	69.69 (13.51; 48.00–90.00)
	Social cue (Study 2)	16	12 males 4 females	5.66 (1.46; 3.42–8.33)	5.87 (1.53; 3.42–8.08)	15.12 (0.22; 15.00–16.50)	65.63 (13.04; 48.00–95.00)
	Non-social cue (Study 2)	16	9 males 7 females	5.13 (1.10; 3.25–7.00)	5.71 (1.34; 3.42–7.50)	15.03 (0.13; 15.00–15.50)	64.75 (12.61; 47.00–88.00)
ASD	No cue (Study 1)	15	13 males 2 females	8.78 (2.92; 5.00–15.58)	5.35 (2.17; 2.75–9.58)	36.43 (6.42; 28.50–48.00)	70.93 (18.89; 40.00–101.00)
	Social cue (Study 2)	17	13 males 4 females	8.60 (1.96; 4.83–11.83)	5.48 (1.79; 2.83–8.08)	35.59 (7.00; 27.00–47.50)	68.71 (17.65; 41.00–113.00)
	Non-social cue (Study 2)	16	13 males 3 females	8.72 (2.83; 4.33–15.42)	5.14 (1.95; 2.92–8.17)	36.50 (7.16; 27.50–58.00)	73.56 (23.83; 47.00–123.00)

Note. TD: typically developing; ASD: autism spectrum disorder; BPVS: British Picture Vocabulary Scale; CARS: childhood autism rating scale. Participants only experienced one condition.

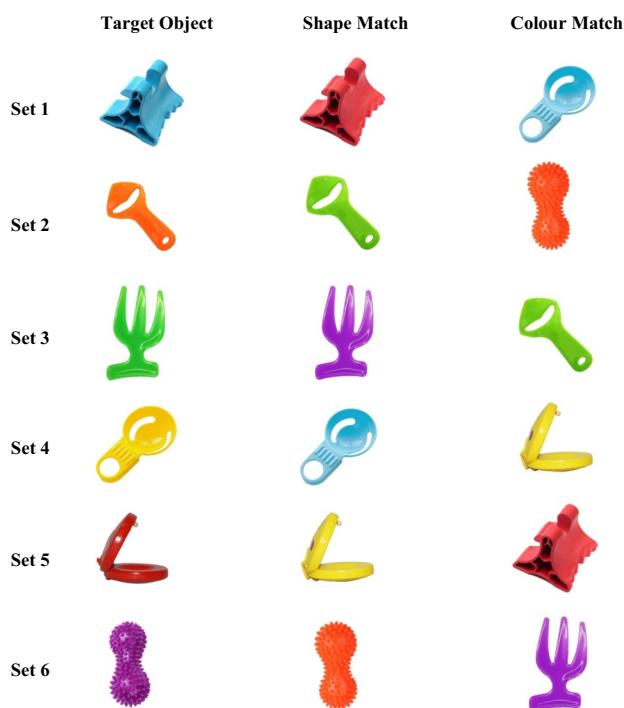


Fig. 1. Sets of unfamiliar objects used in Studies 1 and 2.

2.1.3. Procedure

Participants were tested individually in their own schools and were accompanied by a familiar adult when required. Children were administered the BPVS and Leiter-3 over a series of sessions before completing the word learning task after a one- or two-week break. The word learning task consisted of the following stages delivered in a fixed order: 1. Warm-up trials, 2. Training trials, 3. Delay, 4. Retention and generalisation trials. The word learning task was delivered on the tablet computer, which was placed in front of the child on a table. The experimenter sat quietly while the participant was engaged in the task and offered verbal praise for attention and good behaviour.

2.1.3.1. Warm-up trials. To familiarise children with the format of the task and its response method, children completed four warm-up trials. Children were presented with two pictures of familiar objects positioned to the left and right of the tablet's screen, and heard a female voice (presented through the tablet's speakers) say "Look!" in an excited tone. After viewing the pictures for 3 s, the same voice asked them to identify one of the objects ("Which is the cat? Touch the cat"). Following a correct response, the experimenter issued praise and reinforced the identity of the object (e.g. "Great job, that is the cat!"). Following an incorrect response, the experimenter provided corrective feedback (e.g. "Actually, this is the cat. Can you touch the cat? Well done, you touched the cat!"). The requested object was positioned to the left on two trials, and the right on two trials. Trial order was randomised for each participant. Participants immediately progressed to the first training trial after completing the final warm-up trial.

2.1.3.2. Training trials. This stage provided a context in which children could actively learn six novel word-object pairings. Children completed 48 training trials, administered in two blocks of 24 trials (order counterbalanced across participants). The length of this training stage was based on Smith and Yu's (2008) successful use of 30 trials with TD 12-month-olds. To minimise the likelihood of fatigue, children received a 30-second break after completing the first block of training trials.

On each training trial, children were presented with photographs of two unfamiliar objects positioned to the left and right of the tablet's

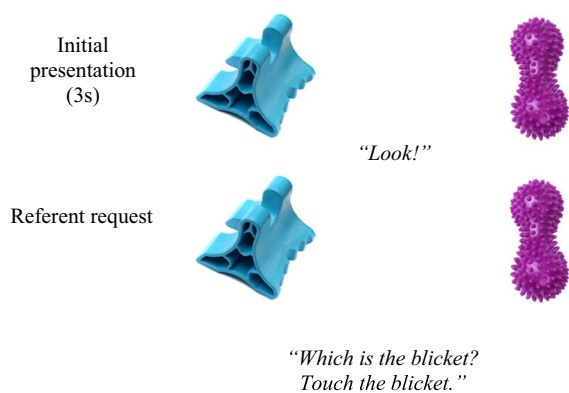


Fig. 2. Example training trial in Study 1.

screen accompanied by a female voice directing them to "Look!" (see Fig. 2). After viewing the pictures for 3 s, the same voice asked them to identify one of the objects (e.g. "Which is the blicket? Touch the blicket"). Following the child's response, the pictures disappeared and the next trial was presented. Individually, these trials are ambiguous (either object could be the word's referent), but cross-situational statistics enabled children to disambiguate the correct word-referent pairings. For example, on the first trial the child might hear the word "blicket" while viewing an X and a Y. At this point, either X or Y could be the target referent. Then, on a subsequent trial, the child experiences "blicket" paired with a Z and an X. If the child is sensitive to cross-trial co-occurrence frequencies, they should correctly map "blicket" onto the X – the object it is experienced with consistently. If a child showed reluctance to respond due to uncertainty, the experimenter encouraged them to guess.

Words occurred an equal number of times in each block and word-object pairings were randomised for each participant. In each block, every object occurred as a target four times and as a foil four times (thus, each target word-object pairing was experienced eight times in total). The target object appeared equally often on the left and right of the screen. The order of trials within blocks was randomised. Children's active learning was tracked by recording the accuracy of their selections. If a child refused to respond on a particular trial, the experimenter could progress to the next trial without selecting a referent object on their behalf by pressing a discrete asterisk button in the upper right-hand corner of the screen. However, the experimenter did not need to use this function in either Study 1 or Study 2.

2.1.3.3. Delay. After completing the last training trial, the child played with the experimenter for 5 min. The tablet was not used during this period.

2.1.3.4. Retention and generalisation trials. To re-engage children's attention to the task, three new warm-up trials with pictures of familiar objects were administered. Here children were presented with three objects per trial positioned to the left, middle, and right of the screen. The requested object appeared in each of the three possible locations on one trial, and trial order was randomised. Thus, children were familiarised with the new "middle" location and selected targets in each location before the retention and generalisation trials.

Children then completed 12 retention trials and 12 generalisation trials (see Fig. 3). Each word was tested on two retention trials and two generalisation trials. For retention trials, three of the named objects were presented on the tablet's screen (positioned to the left, middle, and right). After viewing the pictures for 3 s, children heard a female voice asking them to identify one of the objects (e.g. "Which is the blicket? Touch the blicket"). The purpose of these trials was to assess children's memory of the exact word-referent pairings that were experienced during the training trials. For generalisation trials, children were

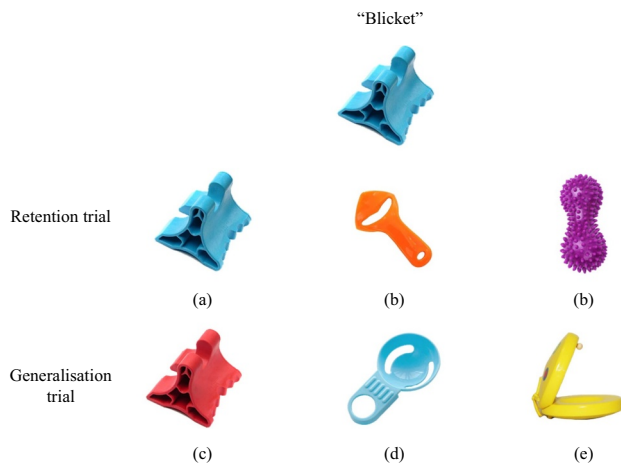


Fig. 3. Example retention and generalisation trials in Studies 1 and 2; (a) target object, (b) non-target objects, (c) shape match for target object, (d) colour match for target object, (e) shape/colour match for non-target object.

presented with differently-coloured variants of named objects (i.e. shape/colour matches). After viewing the pictures for 3 s, children heard a female voice asking them to identify one of the objects (e.g. “Which is the blicket? Touch the blicket”). Crucially, one object was the same shape as the label’s original referent, but different in colour (shape match), while another object matched the original referent on colour, but not shape (colour match). The third object, a shape match for a different named object, served as a distracter. The purpose of these trials was to assess whether children’s extension of labels to novel referents is systematically influenced by shape.

Importantly, all choice objects were of equal familiarity in both retention trials (all had been named in the training trials) and generalisation trials (all were novel objects, although their shapes and colours were introduced in the training trials). Each object was used as a foil on two retention trials and two generalisation trials. To provide the necessary level of control when presenting stimuli, object groupings were fixed. The order of trials within blocks was pseudo-randomised such that the same word was not tested on consecutive trials (e.g. “blicket” retention trial followed by “blicket” generalisation trial), and no more than two trials of the same type occurred consecutively. The target object appeared equally often in each screen location, and never occurred in the same location on more than two consecutive trials. Pictures of objects were exactly the same size as those presented during the training trials. The accuracy and speed of children’s responses were recorded.

2.2. Results

The data for Study 1 and Study 2 are archived at <http://reshare.ukdataservice.ac.uk/853520/>.

2.2.1. Training accuracy

Children’s responses on individual training trials were scored as correct (1) or incorrect (0). The likelihood of responding correctly by chance was 50%. Descriptive statistics for training accuracy are presented in Fig. 4. We investigated the influence of Population and Block on children’s training trial accuracy by conducting a series of generalised linear mixed-effects models (GLMM). The analysis modelled the probability (log odds) of children’s response (correct/incorrect) and all models were conducted using the glmer function from the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). Block was contrast coded as -0.5 (block 1) and $+0.5$ (block 2). Population was coded as 0 (TD) and 1 (ASD).

We started with a baseline model containing by-participant random

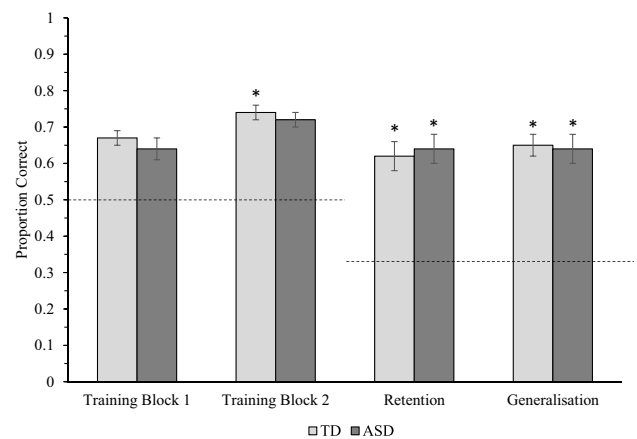


Fig. 4. Mean training, retention, and generalisation trial accuracy for typically developing (TD) children and children with autism spectrum disorder (ASD) in Study 1. Error bars show ± 1 SE. Stars above columns indicate where performance was significantly different from chance, indicated by the dotted lines ($*p < .05$).

intercepts and slopes for Block and by-word random intercepts and slopes for Population x Block (note that Population was a between-subjects variable and so cannot be entered as a random slope by-participant; see Barr, Levy, Scheepers, & Tilly, 2013). Adding Population as a fixed effect did not improve fit in comparison to the baseline model ($p = 1.00$). The addition of Block yielded a significant improvement in fit ($\chi^2 = 7.08, p = .008$). The inclusion of both Block and Population ($p = .77$) or the Block x Population interaction ($p = .96$) did not further improve fit. Therefore, a model containing only Block as a fixed effect provided the best fit to the observed data (see Table 2). These results suggest that both children with ASD and TD children responded with significantly greater accuracy on training trials in Block 2 than Block 1. The populations did not differ significantly in their ability to map novel word-referent relationships based purely on cross-situational statistics.

In order to test whether there was evidence for no difference between the children with ASD and TD children on training accuracy, we determined the Bayes Factor (BF) from the difference in the Bayesian Information Criterion values for models including and excluding Population as a fixed effect (Wagenmakers, 2007). Using this calculation, BF values > 3 are conventionally taken to provide evidence for the hypothesis that populations are similar rather than distinct in their performance (Raftery, 1995). BF values lower than 0.3 are taken to provide evidence for a difference between the populations. BF values between 0.3 and 3 are indeterminate regarding evidence for similarities or differences between populations. For the effect of Population on training accuracy, $BF = 34.8$, providing strong evidence that the training accuracy of children with ASD and TD children was similar. For the interaction between Block and Population, $BF = 38.5$, indicating strong evidence for no interaction between these variables.

We also examined whether individual differences in receptive vocabulary, nonverbal intelligence, and chronological age within each population predicted additional variability in training accuracy. Although the populations did not differ on accuracy, it is possible that

Table 2

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of children’s accuracy on training trials in Study 1, predicted by block.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	0.93	0.16	5.99	< 0.001
Block	0.41	0.13	3.09	0.002
	AIC	BIC	logLik	Deviance
	1708.4	1787.4	-839.2	1409

different factors contributed to their successful performance (individuals with ASD can attain “normal” performance on psychological tasks via atypical routes and compensatory strategies; Happé, 1995; Norbury et al., 2010). As such, the following analyses were conducted on data from the ASD and TD groups separately. We started with models that mirrored the final model detailed above containing Block as a fixed effect, with by-participant and by-word random intercepts and slopes for Block.

For children with ASD, the addition of receptive vocabulary (age equivalent in months) significantly improved fit ($\chi^2 = 6.55, p = .01$). Adding nonverbal intelligence (raw score) afforded a borderline significant improvement in fit ($\chi^2 = 3.65, p = .056$). Adding chronological age did not significantly improve fit ($p = .49$). Including both receptive vocabulary and nonverbal intelligence as fixed effects yielded a borderline significant improvement over the model containing only nonverbal intelligence ($\chi^2 = 3.70, p = .055$), but did not differ from the model containing only receptive vocabulary ($p = .37$). These comparisons indicate that the training accuracy of children with ASD was predicted by receptive vocabulary in addition to Block – children with larger vocabularies were more likely to respond correctly (see Table 3).

For TD children, the addition of receptive vocabulary ($p = .83$), nonverbal intelligence ($p = .49$), or chronological age ($p = 1.00$) did not improve predictive fit when compared with the model containing only Block (which was a significant effect; $Z = 1.99, p = .047$). Thus, the accuracy of TD children on training trials was not reliably influenced by variability in the measured participant traits. Additional models confirmed that population \times receptive vocabulary ($p = .15, BF = 12.8$), population \times nonverbal intelligence ($p = .13, BF = 14.2$), and population \times chronological age ($p = .91, BF = 36.6$) interactions were non-significant, suggesting that individual differences in these abilities did not predict different patterns of performance in training accuracy across groups, with the Bayes Factor values highlighting evidence for no interaction.

For additional analyses investigating relations between response times and accuracy (for Study 1 and Study 2), please see Supplementary materials.

2.2.2. Training response times

Children's response times for correctly-answered training trials were analysed using linear mixed-effects models.² We calculated the average correct response time for each population, in each block. We removed outliers that were over three standard deviations above the mean for the sub-group (e.g. children with ASD in block 1) to rule out trials on which participants were likely distracted. For children with ASD, 479 (97.76%) correct responses were included in our analyses. For TD children, 529 of 541 (97.78%) correct responses were included. With outliers excluded, the mean correct response times for each population are reported in Fig. 5.

We started with a baseline model containing by-participant random intercepts and slopes for Block and by-word random intercepts and slopes for Population \times Block, however, it failed to converge. The baseline model, and all subsequent models, converged when the random effects structure was simplified to by-participant intercepts and slopes for Block and by-word intercepts for Block and Population as main effects (rather than their interaction). Adding Block as a fixed effect did not improve fit in comparison to the baseline model ($p = .65$). Adding population yielded a borderline significant improvement in fit ($\chi^2 = 3.74, p = .053$). Models including both Block and Population ($p = .14, BF = 134.3$) as main effects or the Block \times

² We ran the response time analyses both with log-transformed response times and with a gamma distribution model fit to response times (Lo & Andrews, 2015), and the results were similar in each case. We therefore report analyses on raw response times in the remainder of the paper.

Table 3

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of individual differences in training trial accuracy for children with autism in Study 1, predicted by block and receptive vocabulary.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	-0.43	0.49	-0.89	0.38
Block	0.43	0.22	1.97	0.049
Receptive vocabulary	0.02	0.007	2.97	0.003
AIC		BIC	logLik	Deviance
	785.4	825.8	-383.7	767.4

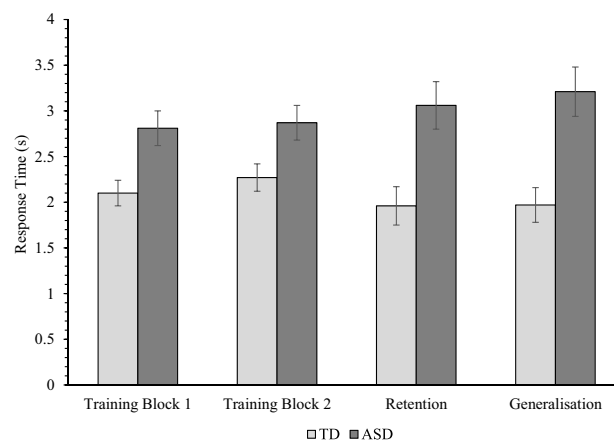


Fig. 5. Mean response times on correctly answered training, retention, and generalisation trials for typically developing (TD) children and children with autism spectrum disorder (ASD) in Study 1. Error bars show ± 1 SE.

Population interaction ($p = .25, BF = 3827.6$) did not differ from the baseline model. These comparisons suggest that a model containing only Population as a fixed factor provides the best fit to the data, suggesting that children with ASD were slower to indicate the correct referent on training trials than TD children (see Table 4).

Our next series of models investigated whether individual differences in receptive vocabulary, nonverbal intelligence, and chronological age predicted additional variability in children's correct response times. Mirroring our analyses for individual differences in training accuracy, these models were conducted on data from the ASD and TD groups separately (so Population, the only fixed effect in the preceding model, was not included as a random slope). Please refer to Supplementary materials for full details of the model building processes for these analyses.

For children with ASD, a model containing chronological age, receptive vocabulary and nonverbal intelligence as fixed effects provided the best fit to the observed data (see Table 5). These results indicate that children with ASD who were older and had higher receptive vocabulary scores were quicker to select the correct referent on training trials. However, higher nonverbal intelligence was associated with slower correct responses.

For TD children, a model containing only chronological age as a fixed effect provided the best fit to the observed data (see Table 6). These results indicate that TD children who were older were quicker to select the correct referent on training trials. However, it is noteworthy that chronological age was correlated with receptive vocabulary ($R = 0.87, p < .001$) and raw nonverbal intelligence score ($R = 0.82, p < .001$), so effects of age should be interpreted cautiously and with these variables in mind.

2.2.3. Retention & generalisation accuracy

Children's responses on individual retention and generalisation trials were scored as correct (1) or incorrect (0). The likelihood of responding correctly by chance was 33%. Descriptive statistics for

Table 4

Summary of the fixed effects in the final linear mixed-effects model of children's response times on correctly-answered training trials in Study 1, predicted by population.

Fixed effects	Estimated coefficient	Std. error	<i>t</i>	Pr(> <i>z</i>)
(Intercept)	2.16	0.32	6.70	< 0.001
Population	0.90	0.45	2.00	0.055
	AIC	BIC	logLik	Deviance
	4501.3	4559.7	-2238.7	4477.3

Table 5

Summary of the fixed effects in the final linear mixed-effects model of individual differences in response times on correctly-answered training trials for children with autism in Study 1, predicted by chronological age, receptive vocabulary, and nonverbal intelligence.

Fixed effects	Estimated coefficient	Std. error	<i>t</i>	Pr(> <i>z</i>)
(Intercept)	5.38	1.33	4.06	0.001
Chronological age	-0.03	0.01	-2.95	0.011
Receptive vocabulary	-0.04	0.01	-2.97	0.010
Nonverbal intelligence	0.04	0.02	2.15	0.049
	AIC	BIC	logLik	Deviance
	2073.0	2101.4	-1029.5	2059.0

Table 6

Summary of the fixed effects in the final linear mixed-effects model of individual differences in response times on correctly-answered training trials for typically developing children in Study 1, predicted by chronological age.

Fixed effects	Estimated coefficient	Std. error	<i>t</i>	Pr(> <i>z</i>)
(Intercept)	4.89	0.72	6.79	< 0.001
Chronological age	-0.04	0.01	-3.87	0.001
	AIC	BIC	logLik	Deviance
	2395.3	2416.6	-1192.6	2385.3

retention and generalisation accuracy are presented in Fig. 4.

We conducted a series of GLMMs to explore the effects of Population, Trial Type, and Training Accuracy. Population was coded as 0 (TD) and 1 (ASD), Trial Type was coded as 0 (retention) and 1 (generalisation), and Training Accuracy was a numerical value between 0 and 24 corresponding to the number of correct responses made by the child in the second block of training trials. Please refer to Supplementary materials for full details of the model building process for this analysis.

A model containing only Training Accuracy as a fixed effect provided the best fit to the observed data (see Table 7). These results indicate that children who achieved greater accuracy in Block 2 of the training trials tended to respond correctly on more retention and generalisation trials. No differences between populations or trial types were detected.

As for training accuracy, we explored whether individual differences within each population predicted additional variability in retention and generalisation trial accuracy. The following analyses were conducted on data from the ASD and TD groups separately. We started

Table 7

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of children's accuracy on retention and generalisation trials in Study 1, predicted by training accuracy.

Fixed effects	Estimated coefficient	Std. error	<i>Z</i>	Pr(> <i>z</i>)
(Intercept)	-3.25	0.54	-6.02	< 0.001
Training accuracy	0.23	0.03	6.64	< 0.001
	AIC	BIC	logLik	Deviance
	911.7	1133.1	-407.8	815.7

with models that mirrored the final model detailed above containing Training Accuracy as a fixed effect, with by-participant and by-word intercepts and slopes for Training Accuracy x Trial Type.

For children with ASD, including receptive vocabulary ($p = .30$), nonverbal intelligence ($p = .57$), or chronological age ($p = 1.00$) did not significantly improve predictive fit when added to Training Accuracy (which was a significant effect; $Z = 5.17$, $p < .001$). The same was true for TD children – adding receptive vocabulary ($p = 1.00$), nonverbal intelligence ($p = .24$), or chronological age ($p = 1.00$) did not significantly improve fit when added to Training Accuracy (a significant effect; $Z = 3.01$, $p = .003$). This finding is consistent with Vlach and DeBrook's (2019) study on receptive vocabulary and retention of learning in TD children. These results suggest that accuracy on retention and generalisation trials was not reliably influenced by receptive vocabulary, nonverbal intelligence, or chronological age in either group. Additional models confirmed that population x receptive vocabulary ($p = .78$, $BF = 5.2$), population x nonverbal intelligence ($p = .41$, $BF = 10.0$), and population x chronological age ($p = .95$, $BF = 66.7$) interactions were not significant.

2.2.4. Retention & generalisation response times

Children's response times for correctly-answered retention and generalisation trials were examined using linear mixed-effects models. Outliers were identified and removed in the same way as for the training trials, and the mean correct response times for each population are reported in Fig. 5. For children with ASD, 228 of 232 (98.28%) correct responses were included in our analyses. For TD children, 238 of 245 (97.14%) correct responses were included.

We started with a baseline model containing by-participant random intercepts and slopes for Trial Type and by-word random intercepts and slopes for Population x Trial Type, however, it failed to converge. The baseline model, and all subsequent models, converged when the random effects structure was simplified to by-participant intercepts and slopes for Trial Type and by-word intercepts for Population. The addition of Population ($t = 2.28$, $p = .03$) significantly improved fit ($\chi^2 = 4.80$, $p = .029$). Adding Trial Type did not improve fit ($p = .84$). The inclusion of both Population and Trial Type ($p = .81$, $BF = 21.1$) or the Population x Trial Type interaction ($p = .97$, $BF = 1200.6$) did not further improve fit. Thus, a model containing only Population as a fixed effect provided the best fit to the observed data (see Table 8), indicating that children with ASD took longer to generate correct responses on retention and generalisation trials than TD children.

Finally, we investigated whether individual differences in receptive vocabulary, nonverbal intelligence, and chronological age predicted additional variability in children's correct response times for retention and generalisation trials. These models were conducted on data from the ASD and TD groups separately. Please refer to Supplementary materials for full details of the model building processes for these analyses.

For children with ASD, the observed data were best predicted by a baseline model with a significant intercept ($t = 7.32$, $p < .001$). The inclusion of fixed effects did not improve fit. These results suggest that correct response times for children with ASD on retention and generalisation trials were not significantly influenced by receptive vocabulary, nonverbal intelligence, or chronological age.

Table 8

Summary of the fixed effects in the final linear mixed-effects model of children's response times on correctly-answered retention and generalisation trials in Study 1, predicted by population.

Fixed effects	Estimated coefficient	Std. error	<i>t</i>	Pr(> <i>z</i>)
(Intercept)	2.02	0.34	5.95	< 0.001
Population	1.10	0.48	2.28	0.03
	AIC	BIC	logLik	Deviance
	2137.7	2175.0	-1059.8	2119.7

Table 9

Summary of the fixed effects in the final linear mixed-effects model of individual differences in response times on correctly-answered retention and generalisation trials for typically developing children in Study 1, predicted by chronological age.

Fixed effects	Estimated coefficient	Std. error	<i>t</i>	Pr(> z)
(Intercept)	5.00	1.04	4.81	< 0.001
Chronological age	−0.05	0.02	−2.94	0.009
	AIC	BIC	logLik	Deviance
	1033.1	1050.4	−511.5	1023.1

For TD children, a model containing only chronological age as a fixed effect provided the best fit to the observed data (see Table 9). These results indicate that TD children who were older were quicker to select the correct referent on retention and generalisation trials.

2.3. Discussion

Study 1 investigated cross-situational word learning as an integrated system, exploring the relationship between mapping, retention, and generalisation of novel labels in TD children and children with ASD. Both populations achieved superior accuracy on training trials in Block 2 than Block 1, demonstrating their ability to overcome referential ambiguity based purely on statistical co-occurrences between words and objects. Participants were clearly sensitive to the gradual accrual of associative information as training proceeded. Training trial accuracy was predicted by receptive vocabulary in children with ASD, but not TD children. Impressively, both populations achieved 60–65% accuracy on delayed retention and generalisation trials, significantly exceeding the chance rate of 33%. Retention and generalisation accuracy was predicted by children's accuracy in Block 2 of the training trials, but was not predicted by individual differences in either population. While the populations did not differ on training or test accuracy, children with ASD were slower to indicate correct referents.

The strong performance of children with ASD on the training trials suggests that cross-situational word mapping accuracy is unimpaired (at a group level) when samples are controlled for language comprehension. This finding is congruent with prior evidence that children with ASD can employ statistical learning mechanisms to correctly derive word meanings (McGregor et al., 2013; Venker, 2019), segment speech (e.g. Haebig et al., 2017), and process visual stimuli (e.g. Foti et al., 2015). As in McGregor et al. (2013), we observed that cross-situational mapping in the ASD group was predicted by receptive vocabulary, but not nonverbal intelligence (once receptive vocabulary was accounted for). This finding may indicate that the ability to disambiguate word meanings using statistical information is strongly related to vocabulary development in ASD. It also suggests that children with ASD who have profoundly impaired language acquisition may have deficits in cross-situational word learning. Consistent with previous studies, TD children's cross-situational word learning accuracy was not significantly related to receptive vocabulary (e.g. Vlach & DeBrook, 2019).

An important feature of Study 1 is that we included tests of retention and generalisation to explore how cross-situational mapping supports longer-term word learning as compared to immediate tests of recognition. In line with our predictions, TD children and children with ASD performed very similarly and with good accuracy on both trial types. Children's accurate recollection of word-referent pairings after a 5-minute delay indicates a link between cross-situational *mapping* and longer-term *learning*. It is well-documented that TD children often forget word-referent pairings that are acquired through fast mapping after 5 min. For example, after achieving 91–100% accuracy on Hartley et al.'s (2019) fast mapping task (study 1), children with ASD responded correctly on ~46% of retention and generalisation trials while TD children responded correctly on ~39% of retention and generalisation

trials. This phenomenon indicates that correct identification of meaning in the context of a naming event is not sufficient to guarantee retention, suggesting that these two aspects of word learning are supported by distinct mechanisms (Horst & Samuelson, 2008). McMurray et al.'s (2012) dynamic associative account posits that word learning requires both 'fast' processes that facilitate disambiguation of referential meaning and 'slow' processes that enable storage and retrieval of lexical representations. According to this theory, slow learning involves children gradually accumulating knowledge of how words map onto objects or actions in their environment over multiple learning instances. Our data are consistent with this theory of slow learning driven by associative learning from cross-situational statistics – both groups responded correctly on 62–65% of retention and generalisation trials, with learning being predicted by cross-situational training accuracy.

However, response times for referent selection were marginally different between the groups during training and were significantly different for the retention and generalisation test trials. Children with ASD were slower than TD children for both tasks. Slow learning, in McMurray et al.'s (2012) model, refers to the gradual accumulation of information about word-referent pairings over multiple trials. However, for this to occur in natural language learning, children must quickly access words and process the potential referents in the environment across fast-moving situations (Christiansen & Chater, 2016). If the child's processing of this information is slowed (as reflected in response times) then the child may be disadvantaged in slow learning of words. We return to this issue in the General discussion.

Still, the ability of both groups to acquire, retain, and generalise word-referent mappings suggests that repeatedly overcoming the challenge of referential ambiguity elicited the formation of robust relationships that may be less susceptible to decay than representations established through fast mapping (cf. Axelsson et al., 2012; Horst & Samuelson, 2008). However, it is unclear whether our participants' retrieval was facilitated by (a) increased opportunities to pair words with referents during training (i.e. greater statistical input compared to fast mapping tasks), or (b) the increased cognitive effort required to track cross-situational associations between words and objects in the absence of ostensive cues or conditions that enable the use of mutual exclusivity (at least from the outset).

Despite substantial delays in language development (relative to chronological age), accuracy of cross-situational word-referent mapping and its relationship to slow learning was not qualitatively atypical in the ASD group (mirroring the results of Hartley et al.'s (2019) fast mapping study). At a group level, their accuracies on training, retention, and generalisation trials closely resembled those of TD children matched on receptive vocabulary and raw nonverbal cognitive ability. However, between-population differences may emerge when comparing how children with, and without, ASD utilise attentional cues in the service of cross-situational word learning.

In Study 2, we investigate whether cross-situational word learning improves when referential ambiguity is potentially reduced. Children with ASD and TD children completed variations of our cross-situational word learning task that incorporated social or non-social attentional cues during training. Half of each population received social cues (head turn and gaze shift towards the target) and half received non-social cues (the target was highlighted by a brightly coloured border) that focussed children's attention on the correct referent for each word.

In accordance with MacDonald et al. (2017), we expected that TD children and children with ASD in these conditions would attain superior training accuracy than children in Study 1 (where no cues were provided). However, we expected ASD to have differing impacts on children's performance in the two cue conditions. Several studies have shown that children with ASD and concomitant language impairments are less likely to follow an adult's head turn in comparison to TD controls (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Leekam, Hunnisett, & Moore, 1998; Warreyn, Roeyers, Oelbrandt, & De Groote, 2005). For example, Leekam et al. (1998) reported that 59% of children

with ASD (aged 8 years on average) followed a head turn cue in comparison to 88% of TD 4-year-olds matched on verbal mental age. Thus, we predicted that the training accuracy of children with ASD may be lower than that of TD controls in the social cue condition. However, we anticipated no between-population differences in training accuracy in the non-social cue condition as we had no reason to believe that children with ASD would be any less likely to notice the appearance of the brightly coloured border than TD children.

Regarding retention accuracy, we reasoned that reducing referential ambiguity during training could have different consequences depending on why learning was strong in Study 1. If retention is facilitated by frequently pairing words with their correct referents, children should respond more accurately when attentional cues are provided during training. Alternatively, if retention is facilitated by tracking and disambiguating meaning across multiple exposures, children may respond less accurately on retention and generalisation trials in the cue conditions. The attentional cues serve to reduce referential ambiguity which, in turn, reduces the cognitive effort required to establish correct word-referent relationships. However, this reduction in effort may result in encoding word-referent representations that are less robust and increasingly susceptible to decay. Importantly, this study will reveal whether the provision of attentional cues enhances cross-situational word learning and will highlight whether the benefits of cue types differ across populations.

3. Study 2: the effects of attentional cues on cross-situational word mapping, retention, and generalisation in typical development and autism

3.1. Method

3.1.1. Participants

Participants were 33 children with ASD (26 males, 7 females; M age = 8.66 years; SD = 2.39) recruited from specialist schools, and 32 TD children (21 males, 11 females; M age = 5.40 years; SD = 1.30) recruited from mainstream schools and nurseries. All children had normal or corrected-to-normal colour vision. None of these children participated in Study 1. Children with ASD were previously diagnosed by a qualified educational or clinical psychologist, using standardised instruments (i.e. Autism Diagnostic Observation Scale and Autism Diagnostic Interview – Revised; Lord et al., 1994; Lord et al., 2002) and expert judgement. Diagnoses were confirmed via the Childhood Autism Rating Scale 2 (Schopler et al., 2010), which was completed by each participant's class teacher.

Half of each population were allocated to the 'Social Cue' condition and half were allocated to the 'Non-social Cue' condition. Children with ASD in the Social and Non-social Cue conditions did not significantly differ on chronological age (t = 0.14, p = .89), BPVS age equivalent (t = 0.52, p = .61), non-verbal IQ (t = 1.25, p = .22), Leiter-3 raw score (t = 0.67, p = .51), or CARS score (t = 0.07, p = .95). TD children in the Social and Non-social Cue conditions did not significantly differ on chronological age (t = 1.17, p = .25), BPVS age equivalent (t = 0.31, p = .76), non-verbal IQ (t = 1.77, p = .09), or Leiter-3 raw score (t = 0.19, p = .85). There were no significant between-population differences in BPVS age equivalent (t = 0.41–1.17, p = .25–.68) or Leiter-3 raw score (t = 0.57–1.31, p = .20–.58) across conditions, however, children with ASD were significantly older (t = 3.84–6.24, p < .001) and had significantly higher CARS scores than TD children (t = 11.67–11.98, p < .001).

Importantly, these samples of children with ASD and TD controls had very similar characteristics to the groups in Study 1 (see Table 1), enabling direct comparison across the two studies. Full comparisons between populations and conditions are provided in Appendix A. There were no significant differences in BPVS age equivalent or Leiter-3 raw score between any groups. Samples of children with ASD did not differ on chronological age or CARS score. Samples of TD children did not

differ on chronological age, but the mean CARS score for TD children in Study 1 (M = 15.38) was significantly higher than that of TD children in the Non-social Cue condition of Study 2 (M = 15.03), $t(30)$ = 2.35, p = .026. However, this difference is unlikely to be clinically meaningful given that the lowest possible score on the CARS is 15 and the samples do not differ on any other metric.

Thirty TD children and all of the children with ASD who participated in this study also participated in Hartley et al. (2019). The experimental tasks for that study and this study were administered one week apart (in a random order). The two studies differed in terms of the novel words and objects used, how the experimental tasks were delivered (presenting real objects vs. a tablet computer), and the learning mechanisms being studied (fast mapping vs. cross-situational learning). As such, we were confident that children's responding would not be influenced by interference effects.

3.1.2. Materials

Stimuli were exactly the same novel words and colour photographs of objects (familiar and unfamiliar) as used in Study 1. Two short videos were created depicting a female face looking forward before turning to look left or right. All stimuli were presented via a Microsoft Surface Pro 4 tablet computer.

3.1.3. Procedure

The paradigm was exactly as described for Study 1 with some adjustments to the warm-up trials and training trials due to introducing social or non-social cues. All other stages, and the order of administration, remained the same.

3.1.3.1. Social cue condition. Children completed four warm-up trials followed by 48 training trials, administered in two blocks of 24 trials. The stimuli and format were as previously described, with one exception: a female face appeared in between each pair of objects. On appearance, the face was looking directly forwards, and children heard a female voice directing them to "Look!". After 3 s, the face turned to one of the objects (the target named object) and the same voice asked them to identify one of the objects (e.g. "Which is the blicket? Touch the blicket"; see Fig. 6). Following the child's response, the pictures disappeared and the next trial was presented. Thus, the face's direction of gaze and corresponding head turn provided cues to the referent of each word, effectively reducing the ambiguity of each individual training trial (see Wu, Gopnik, Richardson, & Kirkham, 2010).

3.1.3.2. Non-social cue condition. As for the social cue condition, children completed four warm-up trials followed by 48 training trials (administered in two blocks). On each trial, children were presented with photographs of two objects positioned to the left and right of the tablet's screen accompanied by a female voice saying "Look!". A star shape also appeared between the two objects. We reasoned that children in the social cue condition would look at the face before the head turn, dividing their attention between the three elements on screen. The star appeared in the initial portion of training trials in the non-social condition to divide children's attention in a similar way before experiencing the directional cue. After viewing the pictures for 3 s, the star disappeared, and one of the objects (the target named object) was "highlighted" by a green border and the same voice asked them to identify one of the objects (e.g. "Which is the blicket? Touch the blicket"; see Fig. 6). Following the child's response, the pictures disappeared and the next trial was presented. Here, the green border provided a non-social attentional cue to the referent of each novel word.

3.2. Results

3.2.1. Training accuracy

Children's responses on individual training trials were scored as

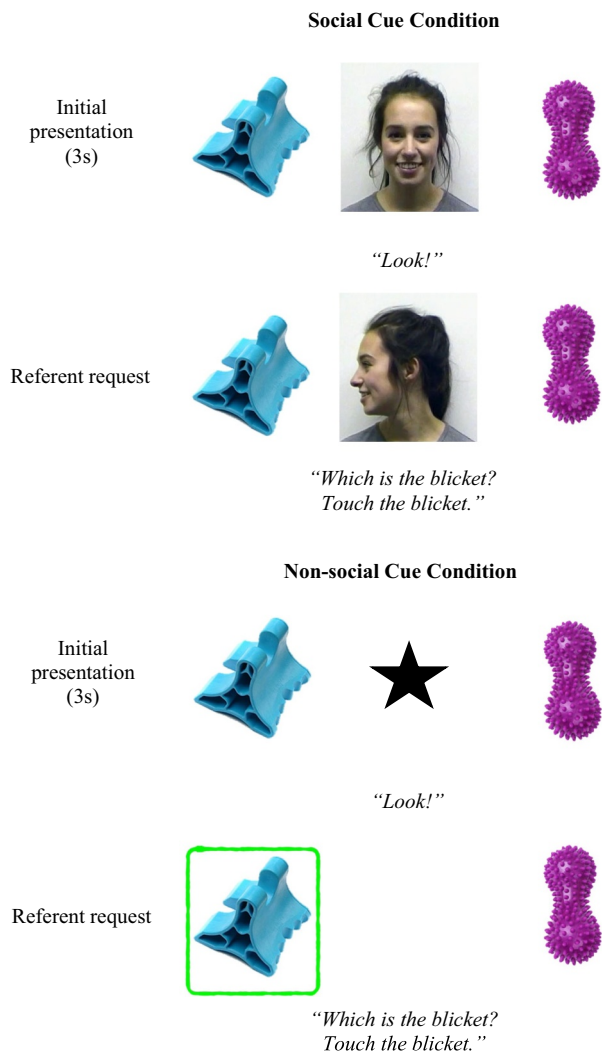


Fig. 6. Example training trials in the social and non-social cue conditions of Study 2.

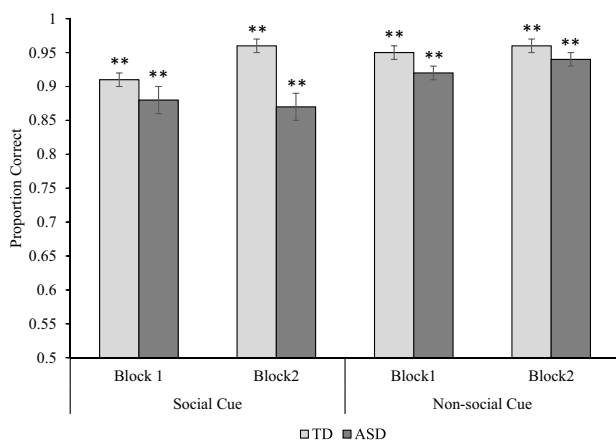


Fig. 7. Mean training trial accuracy for typically developing (TD) children and children with autism spectrum disorder (ASD) in the social cue and non-social cue conditions of Study 2. Error bars show ± 1 SE. Stars above columns indicate where performance was significantly different from chance (0.5; $**p < .001$).

correct (1) or incorrect (0). The likelihood of responding correctly by chance was 50%. Descriptive statistics for training accuracy are presented in Fig. 7.

Please refer to Supplementary materials for full details of the model building processes for all Study 2 analyses. We first explored the effects of Population, Cue Condition, and Block on children's training trial accuracy via a series of GLMMs. Cue Condition was contrast coded as -0.5 (Non-social Condition) and $+0.5$ (Social Condition). Block was contrast coded as -0.5 (block 1) and $+0.5$ (block 2). Population was coded as 0 (TD) and 1 (ASD). The observed data were best predicted by a baseline model with a significant intercept ($Z = 11.30, p < .001$). The inclusion of fixed effects did not improve fit. These results suggest that children's response accuracy on training trials was not reliably influenced by diagnostic group, cue type, or training block. The Bayes Factor results for diagnostic group provided evidence that these groups were similar in accuracy for this task (see Supplementary materials).

Next, we examined whether children's cross-situational mapping was significantly enhanced by the provision of cues by comparing training results from Study 2 and Study 1. Population and Block were coded as described above. Cue Provision was coded 0 (No cue; these data were from children who participated in Study 1) and 1 (Cue provided; these data were from children allocated to the Social and Non-social conditions of Study 2). A model containing fixed effects of Cue Provision and Block provided the best fit to the observed data (see Table 10). These analyses show that both children with ASD and TD children responded with significantly greater accuracy on training trials when cues were provided. They also achieved greater accuracy in Block 2 than Block 1. The Bayes Factor values provided evidence that the diagnostic groups were similar in their accuracies.

As in Study 1, we assessed whether individual differences within each population predicted additional variability in training accuracy in the cue conditions. These models were conducted on data from the ASD and TD groups separately, with the cue conditions collapsed.

For children with ASD, a model containing only nonverbal intelligence as a fixed effect provided the best fit to the observed data (see Table 11). These results indicate that children with ASD with higher nonverbal intelligence were more likely to respond correctly on training trials in the cue conditions. For TD children, receptive vocabulary, nonverbal intelligence, and chronological age did not significantly improve predictive fit over the baseline model.

3.2.2. Training response times

Children's response times for correctly-answered training trials in the Social Cue and Non-social Cue Conditions were examined via linear mixed-effects models. Outliers were identified and removed as described for Study 1. For children with ASD, the analyses included 701 of 714 (98.18%) correct responses in the Social Cue condition and 704 of 712 (98.88%) correct responses in the Non-social Cue condition. For TD children, the analyses included 700 of 718 (97.49%) correct responses in the Social Cue condition and 718 of 731 (98.22%) correct responses in the Non-social Cue condition. With outliers excluded, the mean correct response times for each population are reported in Fig. 8.

A model containing fixed effects of Population and Block provided the best fit to the observed data (see Table 12). These results indicate that children with ASD were slower to respond correctly on training trials (as in Study 1), and both populations were slower to respond correctly in Block 2 than Block 1.

Table 10

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of children's accuracy on training trials in Study 1 and Study 2, predicted by cue provision and block.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	1.03	0.34	3.05	0.002
Cue provision	3.19	0.44	7.21	< 0.001
Block	0.37	0.13	2.73	0.006
	AIC	BIC	logLik	Deviance
	2814.9	3080.9	-1365.5	2730.9

Table 11
Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of individual differences in training trial accuracy for children with autism in Study 2, predicted by block and nonverbal intelligence.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	-0.17	2.12	-0.08	0.94
Nonverbal intelligence	0.06	0.03	2.02	0.043
AIC		BIC	logLik	Deviance
	599.7	676.0	-284.8	569.7

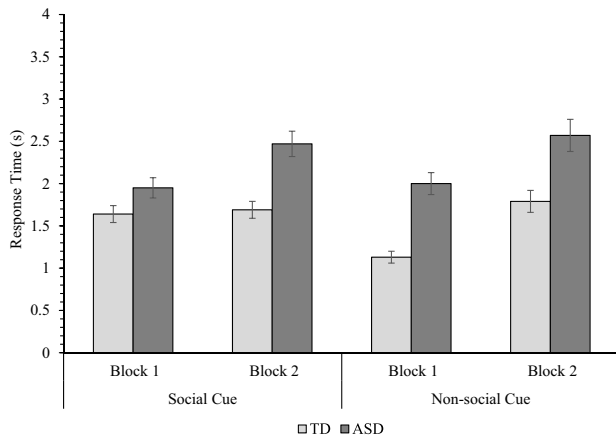


Fig. 8. Mean response times on correctly answered training trials for typically developing (TD) children and children with autism spectrum disorder (ASD) in the Social Cue and Non-social Cue conditions of Study 2. Error bars show ± 1 SE.

Table 12
Summary of the fixed effects in the final general linear mixed-effects model of children's response times on correctly-answered training trials in Study 2, predicted by population and block.

Fixed effects	Estimated coefficient	Std. error	t	Pr(> z)
(Intercept)	1.60	0.17	9.42	< 0.001
Population	0.74	0.24	3.11	0.003
Block	0.47	0.11	4.13	< 0.001
AIC		BIC	logLik	Deviance
	11,021.3	11,096.8	-5497.6	10,995.3

Table 13
Summary of the fixed effects in the final general linear mixed-effects model of children's response times on correctly-answered training trials in Study1 and Study 2, predicted by population, block, and cue provision.

Fixed effects	Estimated coefficient	Std. error	t	Pr(> z)
(Intercept)	2.24	0.24	9.45	< 0.001
Population	0.81	0.22	3.72	< 0.001
Block	0.36	0.10	3.52	< 0.001
Cue provision	-0.69	0.24	-2.85	0.006
AIC		BIC	logLik	Deviance
	15,515.7	15,583.2	-7746.8	15,493.7

We also examined whether children were quicker to generate correct responses on training trials when attentional cues were provided than when no cues were provided in Study 1. A model containing fixed effects of Population, Block, and Cue Provision provided the best fit to the observed data (see Table 13). These results show that Cue Provision significantly predicted variability in children's response times in addition to Population and Block; children in both groups identified correct referents on training trials significantly more quickly in Study 2 than in Study 1.

Table 14
Summary of the fixed effects in the final linear mixed-effects model of individual differences in response times on correctly-answered training trials for children with autism in Study 2, predicted by block and receptive vocabulary.

Fixed effects	Estimated coefficient	Std. error	t	Pr(> z)
(Intercept)	4.00	0.65	6.17	< 0.001
Block	0.62	0.20	3.10	0.005
Receptive vocabulary	-0.03	0.01	-2.66	0.013
AIC		BIC	logLik	Deviance
	5020.2	5069.8	-2500.1	5000.2

As in Study 1, we investigated whether individual differences in receptive vocabulary, nonverbal intelligence, and chronological age predicted additional variability in children's correct response times on training trials in the Social and Non-social Cue conditions. These models were conducted on data from the ASD and TD groups separately.

For children with ASD, a model containing Block and receptive vocabulary as fixed effects provided the best fit to the observed data (see Table 14). These results suggest higher receptive vocabulary was associated with faster correct responding on training trials in the Social and Non-social Cue conditions. For TD children, receptive vocabulary, nonverbal intelligence, and chronological age did not significantly improve predictive fit when added as fixed effects alongside Block (a significant effect, $t = 2.81, p = .009$).

3.2.3. Retention & generalisation accuracy

Children's responses on individual retention and generalisation trials were scored as correct (1) or incorrect (0). The likelihood of responding correctly by chance was 33%. Descriptive statistics for retention and generalisation accuracy are presented in Fig. 9. We analysed the effects of Population, Cue Condition, Trial Type, and Training Accuracy on children's retention and generalisation trial accuracy via a series of GLMMs (all variables coded as described previously).

A model containing Training Accuracy as a fixed effect provided the best fit to the observed data (see Table 15). As in Study 1, children who achieved greater accuracy in Block 2 of the training trials tended to respond correctly on more retention and generalisation trials. No differences between populations, conditions, or trial types were detected (for Population, $BF = 6.8 \times 10^{10}$).

We then explored how children's retention and generalisation accuracy was affected by the provision of attentional cues during mapping in a combined analysis across Studies 1 and 2. A model containing Training Accuracy as a fixed effect provided the best fit to the observed

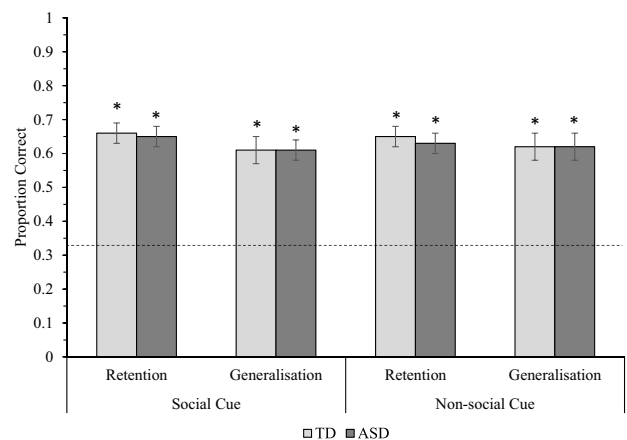


Fig. 9. Mean retention and generalisation trial accuracy for typically developing (TD) children and children with autism spectrum disorder (ASD) in the social cue and non-social cue conditions of Study 2. Error bars show ± 1 SE. Stars above columns indicate where performance was significantly different from chance, indicated by the dotted line ($*p < .05$).

Table 15

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of children's accuracy on retention and generalisation trials in Study 2, predicted by training accuracy.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	-2.80	0.54	-5.24	< 0.001
Training accuracy	0.16	0.03	6.05	< 0.001
	AIC	BIC	logLik	Deviance
	2111.9	2904.0	-907.9	1815.9

Table 16

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of accuracy on retention and generalisation trials in Study 1 and Study 2, predicted by training accuracy.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	-2.22	0.37	-6.05	< 0.001
Training accuracy	0.15	0.02	7.00	< 0.001
	AIC	BIC	logLik	Deviance
	2939.7	3789.5	-1321.8	2643.7

data (see Table 16). Provision of attentional cues during training did not reliably influence retention or generalisation accuracy and did not interact with Population or Trial Type.

As for training accuracy, we assessed whether individual differences within each population predicted additional variability in retention and generalisation trial accuracy for children in the cue conditions. These models were conducted on data from the ASD and TD groups separately, with the cue conditions collapsed.

For children with ASD, a model containing Training Accuracy and chronological age as fixed effects provided the best fit to the observed data (see Table 17). Children with ASD were more likely to respond correctly on retention and generalisation trials in the Social and Non-social Cue conditions as their chronological age increased. However, it is worth noting that chronological age was correlated with receptive vocabulary ($R = 0.59, p < .001$) and raw nonverbal intelligence ($R = 0.69, p < .001$), so effects of age should be interpreted with these variables in mind.

For TD children, a model containing Training Accuracy and non-verbal intelligence as fixed effects provided the best fit to the observed data (see Table 18). TD children with higher nonverbal intelligence were more likely to respond correctly on retention and generalisation trials in the Social and Non-social Cue conditions.

3.2.4. Retention & generalisation response times

Children's response times for correctly-answered retention and generalisation trials were examined. Outliers were identified and removed as described for Study 1. For children with ASD, the analyses included 251 of 257 (97.67%) correct responses in the Social Cue condition and 235 of 241 (97.51%) correct responses in the Non-social Cue condition. For TD children, the analyses included 241 of 245 (98.37%) correct responses in the Social Cue condition and 237 of 244 (97.13%) correct responses in the

Table 17

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of individual differences in retention and generalisation accuracy for children with autism in Study 2, predicted by training accuracy and chronological age.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	-3.93	0.87	-4.53	< 0.001
Training accuracy	0.13	0.03	3.88	< 0.001
Chronological age	0.02	0.01	3.18	0.001
	AIC	BIC	logLik	Deviance
	1032.4	1261.5	-467.2	934.4

Table 18

Summary of the fixed effects in the final generalised linear mixed-effects model (log odds) of individual differences in retention and generalisation accuracy for typically developing children in Study 2, predicted by training accuracy and nonverbal intelligence.

Fixed effects	Estimated coefficient	Std. error	Z	Pr(> z)
(Intercept)	-4.54	1.26	-3.62	< 0.001
Training accuracy	0.12	0.05	2.15	0.032
Nonverbal intelligence	0.04	0.02	2.21	0.027
	AIC	BIC	logLik	Deviance
	972.9	1200.5	-437.5	874.9

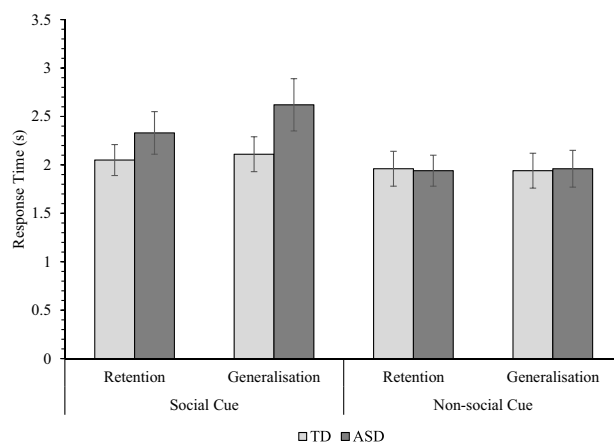


Fig. 10. Mean response times on correctly answered retention and generalisation trials for typically developing (TD) children and children with autism spectrum disorder (ASD) in Study 2. Error bars show ± 1 SE.

Non-social Cue condition. With outliers excluded, the mean correct response times for each population are reported in Fig. 10.

The observed data were best predicted by a baseline model – the inclusion of fixed effects did not improve predictive power. These results suggest that children's response times on correctly-answered trials were not reliably influenced by diagnostic group, type of attentional cue presented during training, or type of trial.

We also tested whether children were faster to generate correct responses on retention and generalisation trials when attentional cues were provided during training than when no cues were provided in Study 1. A model containing Population as a fixed effect provided the best fit to the observed data (see Table 19), suggesting that children with ASD were slower to respond correctly on retention and generalisation trials when data sets for Study 1 and Study 2 were combined.

Finally, we examined whether individual differences in receptive vocabulary, nonverbal intelligence, and chronological age predicted additional variability in children's correct response times on retention and generalisation trials in the Social and Non-social Cue conditions. These models were conducted on data from the ASD and TD groups separately.

For children with ASD, a model containing only receptive vocabulary as a fixed effect provided the best fit to the observed data (see Table 20). These results suggest that children with higher receptive

Table 19

Summary of the fixed effects in the final general linear mixed-effects model of children's response times on correctly-answered retention and generalisation trials in Study1 and Study 2, predicted by population.

Fixed effects	Estimated coefficient	Std. error	t	Pr(> z)
(Intercept)	2.09	0.19	10.91	< 0.001
Population	0.52	0.26	2.03	0.046
	AIC	BIC	logLik	Deviance
	6303.2	6366.4	-3139.6	6279.2

Table 20

Summary of the fixed effects in the final linear mixed-effects model of individual differences in response times on correctly-answered retention and generalisation trials for children with autism in Study 2, predicted by receptive vocabulary.

Fixed effects	Estimated coefficient	Std. error	t	Pr(> z)
(Intercept)	3.82	0.74	5.15	< 0.001
Receptive vocabulary	-0.02	0.01	-2.02	0.052
	AIC	BIC	logLik	Deviance
	2171.0	2191.9	-1080.5	2161.0

Table 21

Summary of the fixed effects in the final linear mixed-effects model of individual differences in response times on correctly-answered retention and generalisation trials for typically developing children in Study 2, predicted by chronological age.

Fixed effects	Estimated coefficient	Std. error	t	Pr(> z)
(Intercept)	3.75	0.58	6.51	< 0.001
Chronological age	-0.03	0.01	-2.97	0.006
	AIC	BIC	logLik	Deviance
	1965.3	1986.1	-977.6	1955.3

vocabulary were quicker to identify correct referents on retention and generalisation trials in the Social and Non-social Cue conditions.

For TD children, a model containing only chronological age as a fixed effect provided the best fit to the observed data (see Table 21). These results suggest that older TD children were quicker to identify correct referents on retention and generalisation trials in the Social and Non-social Cue conditions. However, chronological age was correlated with receptive vocabulary ($R = 0.87, p < .001$) and nonverbal intelligence raw score ($R = 0.82, p < .001$), so it is possible that these variables contributed to the predictive effect.

3.3. Discussion

Experiment 2 investigated whether cross-situational word learning is enhanced by reducing referential ambiguity and whether the benefits of social and non-social attentional cues differ for children with ASD and TD children. Both populations responded with high accuracy on training trials in Block 1 and Block 2 of each cue condition. Between-condition comparisons indicated that training accuracy was not reliably influenced by cue type, population, or block, although children with ASD took significantly longer to identify correct referents. Training accuracy significantly increased with the provision of cues (versus no cues in Study 1) and there was a significant relationship with nonverbal intelligence in children with ASD. Similar to Study 1, all groups achieved 61–66% accuracy on delayed retention and generalisation trials, with performance in both populations predicted by training accuracy, but not cue condition or provision of cues. Additional variability in retention and generalisation accuracy was predicted by chronological age for children with ASD and nonverbal intelligence for TD children.

Unlike children in Study 1, participants in Study 2 could deduce word meanings by simply attending to cues that reliably highlighted the referent for each word. Indeed, our data suggest that both populations may have employed this strategy. TD children responded with ceiling accuracy (91–96%) across conditions and blocks, while children with ASD were marginally less accurate (87–94%) but not significantly different. Both populations appeared to recognise that both social and non-social cues were indicative of referential meaning, despite the lack of instruction explaining this to be the case. Consequently, children reliably directed their attention to correct referents on the vast majority of trials, thus experiencing the almost-maximal frequency of correct word-object pairings during training. When data from Studies 1 and 2 were collapsed to explore the effect of cue provision, Block also

emerged as a main effect across conditions. While the difference between blocks was larger in Study 1, three out of four groups in Study 2 also demonstrated superior accuracy in Block 2 than Block 1. This implies that some participants may have taken a short while to realise the usefulness of attentional cues during training.

Perhaps surprisingly, the training accuracy of TD children and children with ASD in the Social Cue condition did not significantly differ. This demonstrates that our participants with ASD were sensitive to the actor's change in gaze direction and considered this to be a reliable cue to referential meaning. These findings are incongruent with classic evidence that children with ASD do not reliably map word-referent relationships based on gaze (e.g. Baron-Cohen et al., 1997; Preissler & Carey, 2005). However, a growing collection of studies show that children with ASD who develop functional language skills can utilise social-communicative information when identifying the meanings of new words (Bean Ellawadi & McGregor, 2016; Hani, Gonzalez-Barrero, & Nadig, 2013; Luyster & Lord, 2009; McGregor et al., 2013; Norbury et al., 2010). As predicted, the two populations did not differ in their sensitivity to non-social attentional cues. Taken together, data across the two conditions show that our samples of children with ASD utilised attentional cues to overcome the challenge of referential ambiguity when mapping novel words over multiple exposures as effectively as vocabulary-matched TD controls.

Accuracy during training was significantly higher in Study 2 than in Study 1, suggesting that children identified correct referents more reliably when attentional cues were provided. There was a risk that providing highly-accurate attentional cues would trivialise the task of selecting correct referents – if participants were sensitive to the cues then they could correctly select target referents based purely on visual information, without actually learning word-object correspondences. However, the similarity of retention and generalisation accuracy observed in Study 1 and Study 2 suggests that children encoded the novel word-referent correspondences despite the easier learning context. On one hand, this finding suggests that children's exposure to higher frequencies of correct word-referent pairings during training does not necessarily elicit superior “longer term” learning. The fact that TD children's retention accuracy was not promoted by the Social Cue condition also suggests that longer term learning is not necessarily facilitated by opportunities to infer referential intent from social-communicative behaviour. However, on the other hand, our response time analyses show that providing attentional cues can accelerate children's identification of novel word meanings without having a detrimental effect on their subsequent retention or generalisation accuracy. From a practical perspective, this finding has interesting implications for interventions supporting word learning in both typical development and autism.

The similarity of retention and generalisation accuracy across the two populations replicates the results from Study 1, providing further evidence that core relationships between cross-situational mapping and slow learning mechanisms are not atypical in ASD. Both groups established word-referent representations that privileged similarity of shape as the key determinant of referential meaning, enabling accurate extension to unlabelled category members. These findings support recent studies showing that ASD is not characterised by a pervasive deficit in shape-based generalisation (e.g. Field, Allen, & Lewis, 2016; Hartley et al., 2019). However, children with ASD were slower to identify correct referents during training, despite the provision of attentional cues, and when tested after a delay. Whereas the mechanisms of word learning may be intact in children with ASD, the detriment may instead be apparent when learning words under more naturalistic conditions, where the time available for the child to process the input is substantially reduced. Our models incorporating individual differences also suggest that the relative contributions of chronological age, receptive vocabulary, and nonverbal intelligence to mapping and retention may differ for children with ASD and typical development.

4. General discussion

These studies are the first to systematically explore cross-situational word learning as an integrated system of mapping, retention, and

generalisation in both typical development and autism. In comparison to TD controls matched on receptive vocabulary and nonverbal intelligence (raw score), children with ASD were as accurate at disambiguating the meaning of novel words based purely on statistical correspondences. Both groups spontaneously utilised social and non-social attentional cues to facilitate and accelerate their mapping of novel word meanings. Moreover, both groups retained and generalised word-referent representations established through cued and non-cued learning with impressive and comparable accuracy. While children with typical development and autism performed very similarly on measures of learning accuracy, between-group differences emerged in analyses of response times. Children with ASD were significantly slower to correctly identify word meanings than TD children under both cued and non-cued learning conditions. These findings advance understanding of how ASD may affect children's language acquisition and have interesting implications for interventions targeting vocabulary development.

The results of Studies 1 and 2 demonstrate that, under the right conditions, children with ASD can learn novel words as accurately as TD children when expectations are based on current receptive language ability. Our data indicate that ASD does not impair the fundamental mechanisms that underpin cross-situational word learning, or the relationships between them. These findings align with evidence that language acquisition in ASD is not qualitatively atypical (e.g. Boucher, 2012; Ellis Weismer et al., 2011; Gernsbacher, Morson, & Grace, 2015; Gernsbacher & Pripas-Kapit, 2012; Goodwin, Fein, & Naigles, 2012; Naigles, Kelty, Jaffery, & Fein, 2011). Hartley et al. (2019) recently reported that children with ASD (several of whom also participated in the present study) were accurate in their use of mutual exclusivity and retained fast-mapped words with comparable accuracy to TD controls. Moreover, children with ASD who received social feedback after independently mapping word-referent relationships achieved the most accurate retention, outperforming TD controls in the same condition and autistic peers who received non-social feedback or no feedback. Viewed in conjunction with the training data from our Social Cue condition, it appears that – like TD children – children with ASD can benefit from social input both in terms of identifying the meaning of new words and reinforcing longer-term learning.

However, effective word learning performance of children with ASD plus concomitant language delay may be contingent on the quality and pace of input. Our data show that children with ASD were significantly slower to indicate correspondences between words and referents during training and delayed test trials. Typical speech occurs at a rate of approximately 150 words per minute (Studdert-Kennedy, 1986), approximately ten times more rapidly than in the current experimental settings. Slower processing of information in these natural language conditions would result in a reduction of associative strength between words and referents, even if the mechanisms enabling those associations to be formed are unimpaired (McMurray et al., 2012; Yu & Smith, 2012). In the present research, such difficulties that may impact natural language learning were likely mitigated by the highly controlled presentation of visual and auditory stimuli. Firstly, the experiment was delivered via a touchscreen tablet – a platform that children with ASD find extremely engaging (Allen, Hartley, & Cain, 2016). Secondly, the possible referents in our experimental learning environments were substantially constrained – just two options during the training trials, compared to a multitude of possibilities in natural learning situations (Yu & Ballard, 2007; Yurovsky, Smith, & Yu, 2013). Thirdly, children's progression through training and retention trials was regulated by the speed of their responding, meaning their processing time was unrestricted.

The observed profile of good accuracy coupled with slower response times supports recent claims that ASD may delay children's language acquisition by disrupting intake of information, rather than impairing specific word learning processes or strategies per se (Arunachalam & Luyster, 2016, 2018; Hartley, 2019; Hartley et al., 2019; Tenenbaum, Amso, Righi, & Sheinkopf, 2017). That is, skills which appear to be

intact in experimental settings may operate less efficiently in unconstrained natural environments. Evidence from typical development suggests that social cues facilitate word learning in so far as they direct children's attention to referent objects and away from competitors, thereby increasing the accuracy of associative learning mechanisms (Axelsson et al., 2012). However, many children with ASD have diminished social motivation, and may not spontaneously or reliably attend to gaze or gestural cues when deciphering word meanings in real world contexts (Chevallier, Kohls, Troiani, Brodtkin, & Schultz, 2012). Children with ASD may also be overwhelmed by sensory input or distracted by non-relevant aspects of the learning environment (Akechi et al., 2011; Tenenbaum et al., 2017). For example, Akechi et al. (2011) found that children reliably mapped a novel word to a speaker's object (the intended referent) only when it was more perceptually salient than an object they themselves were holding. Thus, advancing understanding of how children with ASD extend experimentally-demonstrated word learning mechanisms to naturalistic settings is an important objective for future research.

For both TD children and children with ASD, word-referent relationships encoded through cross-situational mapping were flexible, enabling extension to novel category members. Our results contrast with previous evidence that shape-based label generalisation is a relative weakness for children with ASD (Hartley & Allen, 2014; Potrzeba et al., 2015; Tek et al., 2008). This discrepancy may be attributable to differences in sample composition and/or methodology.

In terms of population differences, children with ASD may need to acquire a more extensive vocabulary than TD children before they privilege shape as a basis for generalising labels. Previous studies report the absence of a preference for shape-based generalisation in ASD samples with verbal mental ages between 23 and 42 months, whereas the average receptive vocabulary age equivalent of our ASD sample was approximately 64 months. Hartley et al. (2019) also observed a preference for shape-based label extension in children with ASD whose verbal mental ages exceeded 60 months. Viewed alongside those findings, our data support the theory that development of a tendency to generalise labels based on shape is delayed in ASD, but not deviant (Field et al., 2016). In terms of methodology, studies reporting poorer generalisation in ASD usually employ tasks that allow children to select multiple referents for target words (either through explicit gestural responses or through gaze shifting), whereas the present study employed a forced-choice task. It may be that children with ASD prioritise shape when required to select a single referent from an array, but are willing to extend labels based on non-shape properties when several objects can be identified as referents for a target word. Furthermore, unlike studies that present novel word-referent relationships and immediately probe extension (e.g. Hartley & Allen, 2014), we assessed generalisation of already-encoded word-referent relationships after a delay. This approach required children to generalise on the basis of mental representations rather than online categorisation in relation to a visible, or recently visible, target exemplar. The accuracy of our ASD group suggests that their ability to abstract and store mental representations of referents' shapes was not atypical. This, in turn, may suggest that previously reported deficits are linked to categorisation impairments resulting from differences in visual processing and increased attention to an exemplar's category-irrelevant details (e.g. colour; Happé & Frith, 2006). Future research that teases apart online categorisation from prototype formation is required to fully understand the impact of ASD on shape-based generalisation.

Our study also enabled us to test the influence of background individual difference variables in predicting children's responses, and determining whether these background variables had contrasting effects according to group. Though there were effects of age, receptive vocabulary, and nonverbal intelligence relating to accuracy and response times for the children with ASD and TD children across various measures, there was no consistent influence of any of these variables. Furthermore, the intercorrelations between these individual difference

variables mean we cannot be sure that they individually predict variance in performance. Future studies that focus on these individual difference measures, and tease them apart in controlled studies, would help to determine their roles in cross-situational word learning for children with ASD.

From a practical perspective, our findings can potentially inform the development of interventions designed to scaffold word learning in children with ASD. Our study shows that it is possible to teach children new words using a tablet computer. While children with ASD are often highly-motivated to interact with touchscreen technology, evidence of effective learning via this platform is conflicted (see Allen et al., 2016). Here, we demonstrate that touchscreen tablets can be an effective learning platform for children with delayed development, providing that stimuli are presented in a way that appeals to their strengths.

Also, unlike many studies that report word learning deficits in ASD (e.g. Baron-Cohen et al., 1997; Gliga et al., 2012), our tasks required children to actively disambiguate meaning rather than observe naming events. In Study 1, children's mapping of word-referent relationships was completely unguided and, despite the provision of attentional cues in Study 2, children still had to decipher the referent for each word. It may be that children with ASD are more likely to retain word-referent relationships that they actively and independently decode and benefit less when they are passive recipients of ostensive naming (Hartley et al., 2019). This may also be true for TD children, as Zosh and colleagues showed that TD pre-schoolers retained new words with greater accuracy following independent inference of meaning rather than ostensive labelling (Zosh, Brinster, & Halberda, 2013). Thus, in terms of scaffolding longer-term learning, the same hierarchy of effectiveness could apply to both typical development and autism: cross-situational learning may be more effective than mutual exclusivity based fast mapping, which in turn may be more effective than passive ostensive naming. However, this claim is speculative and warrants empirical validation.

Our data also show that children's independent identification of word meanings can be accelerated by providing highly-salient attentional cues, without having a negative impact on their retention accuracy. We observed no differences between cue types, suggesting that any means of guiding children's attention towards a target may be helpful. Although the increased frequency of informative statistical input afforded by attentional cues did not yield a significant improvement in retention after 5 min, the increased speed and accuracy of word-referent mapping may benefit vocabulary development over extended timeframes. It is conceivable that the strategies discussed here could be employed by interventions to optimise language-learning conditions for children with ASD.

Of course, we must address the limitations of this research. It is vital to acknowledge that language development is extremely heterogenous across the ASD spectrum, and deficits in cross-situational word learning accuracy may be observed in minimally verbal samples. It is also important to recognise that our participants with ASD were matched to TD controls on receptive vocabulary, not chronological age. It is well-documented that children with ASD tend to be developmentally delayed on various facets of language development when compared against age norms for TD children (e.g. Charman, Drew, Baird, & Baird, 2003). Hence, we realise that our ASD groups may have attained weaker performance than similarly-aged TD controls. However, in line with other recent work, the purpose of this research was to examine cross-situational word learning across typical and autistic development while controlling for vocabulary. Our samples of TD children and children with ASD were not systematically matched on gender, which could have potentially contributed to between-group differences (although prior work has shown that group differences in gender have little bearing on word learning performance in these populations; e.g. Hartley et al., 2019). In the Social Cue condition of Study 2, we cannot be certain that our participants with ASD would have responded as accurately as TD controls if the actor's shift in gaze was not accompanied by a head turn. It is possible that children's attention was cued towards the target

referent by the head movement, rather than by following the actor's gaze. Future research exploring the use of gaze in the service of cross-situational word learning by children with ASD ought to provide a more stringent test by providing gaze cues without head turns.

Finally, our data only allow us to speculate on potential causes of the slower response times produced by children with ASD. While it is possible that these results are due to differences in language-specific processes, they could alternatively be explained by differences in the generation of motor behaviours. Children with ASD may have processed stimuli at the same speed as TD controls, but generated slower touch responses due to deficits in motor planning and execution (for a review, see Gowen & Hamilton, 2013). Another alternative is that differences in response times indicate that children with ASD were more reflective and less impulsive when selecting referents. However, it is noteworthy that response times tended to increase from block 1 to block 2 during training in Study 2, particularly for children with ASD, which shows that greater practice with the task did not result in faster responding. If children with ASD *require* more time for reflection to accurately learn words, then the bottleneck of natural communicative situations may reduce their performance in comparison to controlled experimental settings, resulting in delayed acquisition.

In summary, our study has elucidated how cross-situational word-referent mapping inter-relates with retention and generalisation in children with ASD and typical development. Despite significant delays in their global language development (relative to chronological age), our ASD samples mapped, retained, and generalised novel words as accurately as TD controls. However, they were significantly slower to generate correct responses across trial types and conditions. These findings imply that fundamental word-learning mechanisms are not atypical in ASD. It may be that ASD affects the efficiency of these mechanisms by disrupting children's intake of language input, particularly during early stages of linguistic development (Arunachalam & Luyster, 2016). Promisingly, we found that providing attentional cues during cross-situational learning can increase both the accuracy and speed of word-referent mapping, without disadvantaging retention in either population. Overall, this research informs understanding of word learning in ASD and spotlights strategies for scaffolding vocabulary acquisition that can be evaluated in future research.

Funding

This research was funded by an Economic and Social Research Council (UK) grant awarded to the first author (ES/N016955/1).

CRediT authorship contribution statement

Calum Hartley: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Laura-Ashleigh Bird:** Methodology, Investigation, Project administration. **Padraic Monaghan:** Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing.

Declaration of competing interest

Hartley declares that he has no conflict of interest. Bird declares that she has no conflict of interest. Monaghan declares that he has no conflict of interest.

Acknowledgements

We would like to thank the children, parents, and schools that supported this research. We would also like to thank Professor Bob McMurray for his guidance on statistical analyses.

Appendix A. Statistical comparisons between groups of typically developing (TD) children and children with autism spectrum disorder (ASD) across Study 1 and Study 2

A.1. Chronological age

	TD no cue (Study 1)	ASD no cue (Study 1)	TD social cue (Study 2)	ASD social cue (Study 2)	TD non-social cue (Study 2)	ASD non-social cue (Study 2)
TD no cue (Study 1)	*	$t = 4.16,$ $p < .001,$ $d = 1.47$	$t = 0.32,$ $p = .75,$ $d = 0.11$	$t = 5.52,$ $p < .001,$ $d = 1.91$	$t = 0.99,$ $p = .33,$ $d = 0.35$	$t = 4.21,$ $p < .001,$ $d = 1.49$
ASD no cue (Study 1)	*	*	$t = 3.80,$ $p = .001,$ $d = 1.35$	$t = 0.20,$ $p = .84,$ $d = 0.07$	$t = 4.66,$ $p < .001,$ $d = 1.65$	$t = 0.05,$ $p = .96,$ $d = 0.02$
TD social cue (Study 2)	*	*	*	$t = 4.88,$ $p < .001,$ $d = 1.70$	$t = 1.17,$ $p = .25,$ $d = 0.41$	$t = 3.84,$ $p = .001,$ $d = 1.36$
ASD social cue (Study 2)	*	*	*	*	$t = 6.24,$ $p < .001,$ $d = 2.18$	$t = 0.14,$ $p = .89,$ $d = 0.05$
TD non-social cue (Study 2)	*	*	*	*	*	$t = 4.73,$ $p < .001,$ $d = 1.67$
ASD non-social cue (Study 2)	*	*	*	*	*	*

A.2. Childhood Autism Rating Scale score

	TD no cue (Study 1)	ASD no cue (Study 1)	TD social cue (Study 2)	ASD social cue (Study 2)	TD non-social cue (Study 2)	ASD non-social cue (Study 2)
TD no cue (Study 1)	*	$t = 13.17,$ $p < .001,$ $d = 4.62$	$t = 0.43,$ $p = .67,$ $d = 0.67$	$t = 15.02,$ $p < .001,$ $d = 4.07$	$t = 2.35,$ $p = .026,$ $d = 0.96$	$t = 11.83,$ $p < .001,$ $d = 4.16$
ASD no cue (Study 1)	*	*	$t = 13.21,$ $p < .001,$ $d = 4.69$	$t = 0.10,$ $p = .92,$ $d = 0.13$	$t = 13.35,$ $p < .001,$ $d = 4.71$	$t = 0.03,$ $p = .98,$ $d = 0.01$
TD social cue (Study 2)	*	*	*	$t = 15.06,$ $p < .001,$ $d = 4.13$	$t = 1.76,$ $p = .088,$ $d = 0.50$	$t = 11.86,$ $p < .001,$ $d = 4.22$
ASD social cue (Study 2)	*	*	*	*	$t = 15.22,$ $p < .001,$ $d = 4.15$	$t = 0.07,$ $p = .95,$ $d = 0.13$
TD non-social cue (Study 2)	*	*	*	*	*	$t = 11.98,$ $p < .001,$ $d = 4.24$
ASD non-social cue (Study 2)	*	*	*	*	*	*

A.3. British Picture Vocabulary Scale (3rd edition) age equivalent

	TD no cue (Study 1)	ASD no cue (Study 1)	TD social cue (Study 2)	ASD social cue (Study 2)	TD non-social cue (Study 2)	ASD non-social cue (Study 2)
TD no cue (Study 1)	*	$t = 0.21,$ $p = .83,$ $d = 0.28$	$t = 0.87,$ $p = .39,$ $d = 0.02$	$t = 0.67,$ $p = .51,$ $d = 0.24$	$t = 1.07,$ $p = .29,$ $d = 0.10$	$t = 1.22,$ $p = .23,$ $d = 0.43$
ASD no cue (Study 1)	*	*	$t = 0.77,$ $p = .45,$ $d = 0.28$	$t = 0.19,$ $p = .85,$ $d = 0.07$	$t = 0.56,$ $p = .58,$ $d = 0.20$	$t = 0.28,$ $p = .78,$ $d = 0.10$
TD social cue (Study 2)	*	*	*	$t = 0.66,$ $p = .51,$ $d = 0.23$	$t = 0.31,$ $p = .76,$ $d = 0.11$	$t = 1.17,$ $p = .25,$ $d = 0.42$
ASD social cue (Study 2)	*	*	*	*	$t = 0.41,$ $p = .68,$ $d = 0.15$	$t = 0.52,$ $p = .61,$ $d = 0.18$
TD non-social cue (Study 2)	*	*	*	*	*	$t = 0.96,$ $p = .35,$ $d = 0.34$
ASD non-social cue (Study 2)	*	*	*	*	*	*

A.4. Leiter-3 nonverbal intelligence test raw score

	TD no cue (Study 1)	ASD no cue (Study 1)	TD social cue (Study 2)	ASD social cue (Study 2)	TD non-social cue (Study 2)	ASD non-social cue (Study 2)
TD no cue (Study 1)	*	$t = 0.78$, $p = .44$, $d = 0.08$	$t = 0.05$, $p = .96$, $d = 0.31$	$t = 0.18$, $p = .86$, $d = 0.06$	$t = 0.29$, $p = .77$, $d = 0.38$	$t = 0.57$, $p = .58$, $d = 0.20$
ASD no cue (Study 1)	*	*	$t = 0.92$, $p = .37$, $d = 0.33$	$t = 0.35$, $p = .73$, $d = 0.12$	$t = 1.08$, $p = .29$, $d = 0.38$	$t = 0.34$, $p = .74$, $d = 0.12$
TD social cue (Study 2)	*	*	*	$t = 0.57$, $p = .58$, $d = 0.20$	$t = 0.19$, $p = .85$, $d = 0.07$	$t = 1.17$, $p = .25$, $d = 0.41$
ASD social cue (Study 2)	*	*	*	*	$t = 0.74$, $p = .47$, $d = 0.26$	$t = 0.67$, $p = .51$, $d = 0.23$
TD non-social cue (Study 2)	*	*	*	*	*	$t = 1.31$, $p = .20$, $d = 0.46$
ASD non-social cue (Study 2)	*	*	*	*	*	*

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2020.104265>.

References

- Akechi, H., Senju, A., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2011). Do children with ASD use referential gaze to learn the name of an object? An eye-tracking study. *Research in Autism Spectrum Disorders*, 5(3), 1230–1242. <https://doi.org/10.1016/j.rasd.2011.01.013>.
- Allen, M. L., Hartley, C., & Cain, K. (2016). iPads and the use of “apps” by children with autism spectrum disorder: Do they promote learning? *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2016.01305>.
- Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., ... Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, 75, 594–604. <https://doi.org/10.1037/0022-006X.75.4.594>.
- Arunachalam, S., & Luyster, R. J. (2016). The integrity of lexical acquisition mechanisms in autism spectrum disorders: A research review. *Autism Research*, 9(8), 810–828. <https://doi.org/10.1002/aur.1590>.
- Arunachalam, S., & Luyster, R. J. (2018). Lexical development in young children with autism spectrum disorder (ASD): How ASD may affect intake from the input. *Journal of Speech, Language, and Hearing Research*, 61(11), 2659–2672. https://doi.org/10.1044/2018_JSLHR-L-RSAUT-18-0024.
- Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3, 1–8. <https://doi.org/10.3389/fpsyg.2012.00088>.
- Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, 29(5), 832–843. <https://doi.org/10.1037/0012-1649.29.5.832>.
- Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, 68, 48–57. <https://doi.org/10.2307/1131924>.
- Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1–7. 2014Institute for Statistics and Mathematics of WU website <http://CRAN.R-project.org/package=lme4>.
- Bavin, E. L., Kidd, E., Prendergast, L. A., & Baker, E. K. (2016). Young children with ASD use lexical and referential information during on-line sentence processing. *Frontiers in Psychology*, 7, 171. <https://doi.org/10.3389/fpsyg.2016.00171>.
- Bean Ellawadi, A., & McGregor, K. K. (2016). Children with ASD can use gaze to map new words. *International Journal of Language & Communication Disorders*, 51(2), 212–218. <https://doi.org/10.1111/1460-6984.12192>.
- Bedford, R., Gliga, T., Frame, K., Hudry, K., Chandler, S., Johnson, M. H., & Charman, T. (2013). Failure to learn from feedback underlies word learning difficulties in toddlers at risk for autism. *Journal of Child Language*, 40(01), 29–46. <https://doi.org/10.1017/S0305000912000086>.
- Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39–53. <https://doi.org/10.1016/j.cognition.2012.08.008>.
- Bloom, P. (2002). Mindreading, communication, and the learning of names for things. *Mind & Language*, 17, 37–54. <https://doi.org/10.1111/1468-0017.00188>.
- Boucher, J. (2012). Research review: Structural language in autistic spectrum disorder - Characteristics and causes. *Journal of Child Psychology and Psychiatry*, 53(3), 219–233. <https://doi.org/10.1111/j.1469-7610.2011.02508.x>.
- Carey, S. (1978). The child as a word learner. In M. Halle, J. Bresnan, & A. Miller (Eds.). *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.
- Cartmill, E., Armstrong, B. F., Gleitman, L. R., Goldin-Meadows, S., Medinac, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110, 11278–11283. <https://doi.org/10.1073/pnas.1309518110>.
- Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language*, 30(01), 213–236. <https://doi.org/10.1017/S0305000902005482>.
- Chen, C. H., Zhang, Y., & Yu, C. (2018). Learning object names at different hierarchical levels using cross-situational statistics. *Cognitive Science*, 42, 591–605. <https://doi.org/10.1111/cogs.12516>.
- Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Science*, 16, 231–239. <https://doi.org/10.1016/j.tics.2012.02.007>.
- Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, 1–72. <https://doi.org/10.1017/S0140525X1500031X>.
- de Marchena, A., Eigsti, I.-M., Worek, A., Ono, K. E., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, 119(1), 96–113. <https://doi.org/10.1016/j.cognition.2010.12.011>.
- Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British picture vocabulary scale* (2nd ed.). Windsor: NFER-Nelson.
- Ellis Weismer, S., Gernsbacher, M. A., Stronach, S., Karasinski, C., Eernisse, E. R., Venker, C. E., & Sindberg, H. (2011). Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. *Journal of Autism and Developmental Disorders*, 41(8), 1065–1075. <https://doi.org/10.1007/s10803-010-1134-4>.
- Fenson, L., Marchman, V. A., Thal, D. J., Reznick, S., & Bates, E. (2006). *The MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore: Brookes.
- Field, C., Allen, M. L., & Lewis, C. (2016). Are children with autism spectrum disorder initially attuned to object function rather than shape for word learning? *Journal of Autism and Developmental Disorders*, 46, 1210–1219. <https://doi.org/10.1007/s10803-015-2657-5>.
- Foti, F., De Crescenzo, F., Vivanti, G., Menghini, D., & Vicari, S. (2015). Implicit learning in individuals with autism spectrum disorders: A meta-analysis. *Psychological Medicine*, 45(5), 897–910. <https://doi.org/10.1017/S0033291714001950>.
- Gernsbacher, M. A., Morson, E. M., & Grace, E. J. (2015). Language development in autism. In G. Hickok, & S. Small (Eds.). *Neurobiology of language* (pp. 879–886). San Diego, CA: Elsevier.
- Gernsbacher, M. A., & Pripas-Kapit, S. (2012). Who's missing the point? A commentary on claims that autistic persons have a specific deficit in figurative language comprehension. *Metaphor and Symbol*, 27, 93–105. <https://doi.org/10.1080/10926488.2012.656255>.
- Gliga, T., Elsabbagh, M., Hudry, K., Charman, T., Johnson, M., & The BASIS Team (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development*, 83, 926–938. <https://doi.org/10.1111/j.1467-8624.2012.01750.x>.
- Goodwin, A., Fein, D., & Naigles, L. R. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. *Autism Research*, 5, 109–123. <https://doi.org/10.1002/aur.1220>.

- Gowen, E., & Hamilton, A. (2013). Motor abilities in autism: A review using a computational context. *Journal of Autism and Developmental Disorders*, 43(2), 323–344. <https://doi.org/10.1007/s10803-012-1574-0>.
- Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 58(11), 1251–1263. <https://doi.org/10.1111/jcpp.12734>.
- Hani, H. B., Gonzalez-Barrero, A. M., & Nadig, A. S. (2013). Children's referential understanding of novel words and parent labeling behaviors: Similarities across children with and without autism spectrum disorders. *Journal of Child Language*, 40(5), 971–1002. <https://doi.org/10.1017/S0305000912000426>.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task: Performance of subjects with autism. *Child Development*, 66, 843–855. <https://doi.org/10.2307/1131954>.
- Happé, F., & Frith, U. (2006). The weak central coherence account: Detail focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36, 5–25. <https://doi.org/10.1007/s10803-005-0039-0>.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.
- Hartley, C. (2019). Language acquisition in children with autism spectrum disorder. In J. Horst, & J. V. K. Torkildsen (Eds.). *International handbook of language development* (pp. 404–424). Routledge.
- Hartley, C., & Allen, M. L. (2014). Generalisation of word-picture relations in children with autism and typically developing children. *Journal of Autism and Developmental Disorders*, 44, 2064–2071. <https://doi.org/10.1007/s10803-014-2074-1>.
- Hartley, C., Bird, L.-A., & Monaghan, P. (2019). Investigating the relationship between fast mapping, retention, and generalisation of words in children with autism spectrum disorder and typical development. *Cognition*, 187, 126–138. <https://doi.org/10.1016/j.cognition.2019.03.001>.
- Henderson, L., Powell, A., Gareth Gaskell, M., & Norbury, C. (2014). Learning and consolidation of new spoken words in autism spectrum disorder. *Developmental Science*, 17(6), 858–871. <https://doi.org/10.1111/desc.12169>.
- Horst, J. S., & Hout, M. C. (2016). The novel object and unusual name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409. <https://doi.org/10.3758/s13428-015-0647-3>.
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13, 128–157. <https://doi.org/10.1080/15250000701795598>.
- Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, 13, 706–713. <https://doi.org/10.1111/j.1467-7687.2009.00926.x>.
- Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, 33(1), 3–13.
- Kalashnikova, M., Mattock, K., & Monaghan, P. (2016). Mutual exclusivity develops as a consequence of abstract rather than particular vocabulary knowledge. *First Language*, 36(5), 451–464. <https://doi.org/10.1177/0142723716648850>.
- Kohnert, K. J., Bates, E., & Hernandez, A. E. (1999). Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research*, 42(6), 1400–1413. <https://doi.org/10.1044/jslhr.4206.1400>.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299–321. [https://doi.org/10.1016/0885-2014\(88\)90014-7](https://doi.org/10.1016/0885-2014(88)90014-7).
- Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology*, 15(1), 77–95. <https://doi.org/10.1111/j.2044-835X.1997.tb00726.x>.
- Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39(7), 951–962. <https://doi.org/10.1111/1469-7610.00398>.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, 1171. <https://doi.org/10.3389/fpsyg.2015.01171>.
- Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2002). *Autism diagnostic observation schedule* (WPS ed.). Los Angeles: Western Psychological Services.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview – Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685. <https://doi.org/10.1007/BF02172145>.
- Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology*, 45(6), 1774–1786. <https://doi.org/10.1037/a0016223>.
- MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, 94, 67–84. <https://doi.org/10.1016/j.cogpsych.2017.02.003>.
- Mayo, J., & Eigsti, I. M. (2012). Brief report: A comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. *Journal of Autism and Developmental Disorders*, 42(11), 2476–2485. <https://doi.org/10.1007/s10803-012-1493-0>.
- McGregor, K. K., Rost, G., Arenas, R., Farris-Trimble, A., & Stiles, D. (2013). Children with ASD can use gaze in support of word recognition and learning. *Journal of Child Psychology and Psychiatry*, 54(7), 745–753. <https://doi.org/10.1111/jcpp.12073>.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119, 831–877. <https://doi.org/10.1037/a0029872>.
- Monaghan, P. (2017). Canalization of language structure from environmental constraints: A computational model of word learning from multiple cues. *Topics in Cognitive Science*, 9, 21–34. <https://doi.org/10.1111/tops.12239>.
- Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition*, 123(1), 133–143. <https://doi.org/10.1016/j.cognition.2011.12.010>.
- Naigles, L. R., Kelty, E., Jaffery, R., & Fein, D. (2011). Abstractness and continuity in the syntactic development of young children with autism. *Autism Research*, 4(6), 422–437. <https://doi.org/10.1002/aur.223>.
- Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, 5(4), 203–234. <https://doi.org/10.1080/15475440903167528>.
- Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: Word learning in autistic disorders. *Neuropsychologia*, 48(14), 4012–4019. <https://doi.org/10.1016/j.neuropsychologia.2010.10.015>.
- Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., ... Kjellmer, L. (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. *Autism*, 19(8), 934–943. <https://doi.org/10.1177/1362361314556782>.
- Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, 7, 1245. <https://doi.org/10.3389/fpsyg.2016.01245>.
- Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child Development*, 78, 1265–1287. <https://doi.org/10.1111/j.1467-8624.2007.01065.x>.
- Pataai, S., & Diesendruck, G. (2008). Intentions help children learn meaningful rules. *Journal of Child Language*, 35, 221–234. <https://doi.org/10.1017/S0305000907008331>.
- Potrzeba, E. R., Fein, D., & Naigles, L. (2015). Investigating the shape bias in typically developing children and children with autism spectrum disorders. *Frontiers in Psychology*, 6, 446. <https://doi.org/10.3389/fpsyg.2015.00446>.
- Preissler, M. A., & Carey, S. (2005). What is the role of intentional inference in word learning? Evidence from autism. *Cognition*, 97, B13–B23. <https://doi.org/10.1016/j.cognition.2005.01.008>.
- Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge, MA: MIT Press.
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.). *Sociological methodology 1995* (pp. 111–196). Cambridge, MA: Blackwell.
- Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language*, 87, 105–127. <https://doi.org/10.1016/j.jml.2015.09.005>.
- Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter international performance scale (Leiter-3)*. Los Angeles: Western Psychological Services.
- Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, 60(5), 464–477. <https://doi.org/10.1111/jir.12284>.
- Roser, M. E., Aslin, R. N., McKenzie, R., Zahra, D., & Fiser, J. (2015). Enhanced visual statistical learning in adults with autism. *Neuropsychology*, 29(2), 163. <https://doi.org/10.1037/neu0000137>.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood autism rating scale-CARS-2* (2nd ed.). Los Angeles: Western Psychological Services.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558–1568. <https://doi.org/10.1016/j.cognition.2007.06.010>.
- Studdert-Kennedy, M. (1986). Some developments in research on language behavior. In N. J. Smelser, & D. R. Gerstein (Eds.). *Behavioral and social science: Fifty years of discovery: In commemoration of the fiftieth anniversary of the "Ogburn report: Recent social trends in the United States"* (pp. 208–248). Washington, DC: National Academy Press.
- Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, 126, 395–411. <https://doi.org/10.1016/j.jecp.2014.06.003>.
- Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning? *Autism Research*, 1, 208–222. <https://doi.org/10.1002/aur.38>.
- Tenenbaum, E. J., Amso, D., Righi, G., & Sheinkopf, S. J. (2017). Attempting to “increase intake from input”: Attention and word learning children with autism. *Journal of Autism and Developmental Disorders*, 47, 1791–1805. <https://doi.org/10.1007/s10803-017-3098-0>.
- Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, 23(1), 157–176. <https://doi.org/10.1017/S0305000900010138>.
- Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66(1), 126–156. <https://doi.org/10.1016/j.cogpsych.2012.10.001>.
- Venker, C. E. (2019). Cross-situational and ostensive word learning in children with and without autism spectrum disorder. *Cognition*, 183, 181–191. <https://doi.org/10.1016/j.cognition.2018.10.025>.
- Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(4), 700–711. <https://doi.org/10.1037/xlm0000611>.
- Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, 127(3), 375–382. <https://doi.org/10.1016/j.cognition.2013.03.001>.

- 2013.02.015.
- Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, 45(6), 1611–1617. <https://doi.org/10.1037/a0016134>.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/BF03194105>.
- Warreyn, P., Roeyers, H., Oelbrandt, T., & De Groote, I. (2005). What are you looking at? Joint attention and visual perspective taking in young children with autism spectrum disorder. *Journal of Developmental and Physical Disabilities*, 17(1), 55–73. <https://doi.org/10.1007/s10882-005-2201-1>.
- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2010). Social cues support learning about objects from statistics in infancy. In S. Ohlsson, & R. Catrambone (Eds.). *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 1228–1233). Austin, TX: Cognitive Science Society.
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70, 2149–2165. <https://doi.org/10.1016/j.neucom.2006.01.034>.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420. <https://doi.org/10.1111/j.1467-9280.2007.01915.x>.
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262. <https://doi.org/10.1016/j.cognition.2012.06.016>.
- Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, 145, 53–62. <https://doi.org/10.1016/j.cognition.2015.07.013>.
- Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, 21(1), 1–22. <https://doi.org/10.3758/s13423-013-0443-y>.
- Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, 16(6), 959–966. <https://doi.org/10.1111/desc.12036>.
- Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science*, 17(1), 20–28. <https://doi.org/10.1080/10888691.2013.748420>.