RESEARCH ARTICLE

Processing of Co-Reference in Autism Spectrum Disorder

Philippa L. Howard ⓘ, Simon P. Liversedge, and Valerie Benson

Accuracy for reading comprehension and inferencing tasks has previously been reported as reduced for individuals with autism spectrum disorder (ASD), relative to typically developing (TD) controls. In this study, we used an eye movements and reading paradigm to examine whether this difference in performance accuracy is underpinned by differences in the inferential work required to compute a co-referential link. Participants read two sentences that contained a category noun (e.g., *bird*) that was preceded by and co-referred to an exemplar that was either typical (e.g., *pigeon*) or atypical (e.g., *penguin*). Both TD and ASD participants showed an effect of typicality for gaze durations upon the category noun, with longer times being observed when the exemplar was atypical, in comparison to typical. No group differences or interactions were detected for target processing, and verbal language proficiency was found to predict general reading and inferential skill. The only difference between groups was that individuals with ASD engaged in more re-reading than TD participants. These data suggest that readers with ASD do not differ in the efficiency with which they compute anaphoric links on-line during reading. ***Autism Res** 2017, 10: 1968–1980.* © 2017 International Society for Autism Research, Wiley Periodicals, Inc.

Lay Summary: Individuals with autism spectrum disorder (ASD) have previously been reported to have difficulties with reading comprehension. This study examined whether a difference in the speed with which individuals with ASD form connections between words (co-reference processing) may contribute to comprehension difficulties. No evidence was found to suggest that ASD readers differ to typically developing readers in the speed of co-reference processing. Therefore, this data would suggest that differences in co-reference processing are unlikely to account for reading comprehension difficulties in ASD.

Keywords: co-reference; discourse; eye movements; reading

Introduction

Autism Spectrum Disorder (ASD) is diagnosed when an individual has significant difficulty with social interaction and communication, in addition to restricted and repetitive patterns of behaviour [American Psychiatric Association, 2013]. Basic reading ability (e.g., word identification) and the efficiency of such processing is found to be intact in individuals with ASD without learning difficulties or language impairment [e.g., Howard, Liversedge, & Benson, 2017a; Huemer & Mann, 2010; Minshew, Goldstein, & Siegel, 1995; Saldaña, Carreiras, & Frith, 2009]. However, ASD is associated with atypical performance for higher-order reading tasks, such as comprehension and inferencing.

In general, comprehension accuracy is reported as reduced for readers with ASD, in comparison to typically developing (TD) controls and what would be predicted based upon IQ [Huemer & Mann, 2010; Jones et al., 2009; Nation, Clarke, Wright, & Williams, 2006; cf.,

Åsberg, Kopp, Berg-Kelly, & Gillberg, 2010]. A meta-analysis indicated that ASD does not independently cause comprehension difficulties, but increases the likelihood of these difficulties occurring [Brown, Oram-Cardy, & Johnson, 2013] that are highly associated with verbal language proficiency [e.g., Lucas & Norbury, 2015; Norbury & Nation, 2011]. Therefore, it would seem that the cognitive processing differences and behavioural features associated with ASD may contribute but are not an autonomous cause of reduced comprehension accuracy.

The findings in relation to inferencing accuracy in ASD are more consistent. For operational purposes, we adopt the broad definition that an inference is any implicit information or link readers draw from text [e.g., Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992]. This therefore includes aspects of referential processing (e.g., identifying that *he* refers to *Dave* when reading *Trevor admired Dave, he had an excellent work ethic*), where a reader infers that words refer to the same semantic entity, in order for a co-referential link

to be formed [e.g., Ehrlich & Rayner, 1983]. Other types of inferential processing include causal or bridging inferences, where events are inferred to link portions of text together [e.g., inferring that it had rained when reading *Sally had forgotten her umbrella, she was soaked when she arrived at work*, e.g., Haviland & Clark, 1974]. Causal inferences are generally considered to be more cognitively effortful than referential processing, because of the requirement to infer more complex information (i.e., causal relations in events). In addition, readers compute global pragmatic inferences [e.g., a character's intent, e.g., Poynor & Morris, 2003] that embellish global text coherence. This might be considered even more complex than both referential and causal inferencing, as it involves generating inferences that are elaborative and relate to global discourse.[1] Inferences have previously been categorized into a hierarchy according to various criteria, for example, whether an inference is automatically or strategically computed [e.g., McKoon & Ratcliff, 1992], and, whether an inference is text-connecting and necessary for local coherence or "extratextual" serving to embellish the global text representation [e.g., Graesser et al., 1994]. Irrespective of the deemed complexity/categorisation of an inference, readers engage in such processing and integrate inferred information into the discourse model, in order to maximise local and global text coherence that is often necessary for proficient comprehension.

There are multiple reports of individuals with ASD performing with reduced accuracy in comparison to TD controls on reading and aural tasks that require an inference to be computed [e.g., Bodner, Engelhardt, Minshew, & Williams, 2015; Dennis, Lazenby, & Lockyer, 2001; Jolliffe & Baron-Cohen, 1999, 2000; Minshew et al., 1995; Norbury & Bishop, 2002; Norbury & Nation, 2011; O'Connor & Klein, 2004; Tirado & Saldaña, 2016; cf., Saldaña & Frith, 2007]. Therefore, it is possible that ASD specific difficulties associated with inferential processing are a contributing factor to reading comprehension difficulties. Jolliffe and Baron-Cohen [1999] found adults diagnosed with autism and Asperger's Syndrome to be less accurate at answering multiple-choice questions about sentences that evoked a bridging inference (see Example 1 below), in comparison to TD controls. Also, individuals with autism took longer to respond in comparison to TD controls and participants with Asperger's syndrome.

1. George left his bath water running. George cleared up the mess in the bathroom.
   Question: George cleared up the mess in the bathroom because:

The bath had overflowed
His brother had left it untidy
The workman hadn't cleared up his mess

Further, there are reports of individuals with ASD having reduced accuracy for comprehension questions that require inferential processing, but intact comprehension when the answer can be derived from information explicitly provided within a text [Jolliffe & Baron-Cohen, 2000; Norbury & Bishop, 2002; Norbury & Nation, 2011]. What these studies suggest is that inferential difficulties in ASD cannot necessarily be attributed to reduced comprehension in general, and that inferencing efficiency (speed) may be associated with the severity of ASD symptoms.

Similarly to what has been found for comprehension, inferencing accuracy in ASD is related to verbal language proficiency [as is also the case for TD individuals, e.g., Norbury & Nation, 2011]. Lucas and Norbury [2015] found vocabulary knowledge and verbal working memory to predict inferencing accuracy over and above ASD status. However, a data trend indicated that a higher proportion of participants with ASD and language impairment (50%), and ASD without language impairment (33.3%) had specific difficulties with inferencing, in comparison to TD controls (12.5%). In addition, ASD has been reported to account for approximately 10% of the variance in inferencing accuracy scores [Norbury & Nation, 2011]. It would therefore seem that both verbal language proficiency and ASD independently impact upon inferencing skill, and therefore that inferencing difficulties in ASD may affect reading comprehension. For example, interventions aimed at assisting adolescents with ASD to compute referential links during reading, have been found to improve comprehension [e.g., O'Connor & Klein, 2004]. However, based upon off-line behavioural studies, it is difficult to identify the nature and time course of mechanistic processing differences that exist in relation to inferential processing in ASD.

A question that remains to be addressed relates to how ASD impacts upon the time-course of on-line inferential processing? One study that attempted to answer this question asked adolescents to read two sentences that evoked a bridging inference [replicating Saldaña & Frith, 2007], as their eye movements were monitored [Sansosti, Was, Rawson, & Remaklus, 2013]. Readers with ASD were reported to have longer fixation durations, make more regressive saccades, and have longer reading times. This suggests that the computation of an inference was more effortful for readers with ASD and the authors concluded that this reflected an atypicality in the integration of world knowledge into the discourse model. However, only global eye movement measures were reported (averaged across entire trials). In order to investigate the precise difference in on-line eye movement behaviour

---

[1]This is not an exhaustive list. There are many other inferences readers compute. For a comprehensive description of these, see Graesser et al. (1994).

during reading, assessment of measures localised to critical regions within sentences is necessary. Nevertheless, Sansosti et al.'s study suggests that the moment-to-moment computation of inferences is atypical in ASD.

More recently, Micai, Joseph, Vulchanova, and Saldaña [2017] asked adolescents with and without ASD who were matched on a range of cognitive variables, including oral language skill, to read passages as their eye movements were monitored. The text required an inference to be formed and the question directly probed participant's computation of this inference. For an example, see 2 below, where participants were expected to infer that Mico was a cat, upon identification of the target word *mouse*.

2. It was Monday morning and was really warm. Mr. Francisco fed his parrot and then went over to check the little Mico was ok. He was in a deep sleep and appeared to be dreaming. Mico's legs were moving back and forth as if he was imagining chasing a mouse very fast, trying to catch it.

    What animal is Mico?
    Dog
    Parrot
    Cat

The two groups did not differ in comprehension accuracy, suggesting that both TD and ASD participants had generated the correct inference. However, ASD participants had longer gaze durations than TD readers upon the critical word that informed the inference (*mouse*). In addition, ASD readers regressed back to words that supported this inference (e.g., *little*), on a higher proportion of trials, in comparison to TD readers. This demonstrates that there are subtle differences in the efficiency of inferential processing during reading in ASD.

Both the Weak Central Coherence (WCC) theory [Frith & Happé, 1994] and theory of Complex Information Processing [CIP; Minshew & Goldstein, 1998] predict integrative processes to be atypical in ASD. However, it would appear that readers with ASD are as efficient as TD readers at constructing a mental representation of what a text explicitly conveys [Au-Yeung, Kaakinen, Liversedge, & Benson, 2015; Howard et al., 2017a]. Therefore, the hypothesis that readers with ASD have a universal difficulty with text integration is not supported. However, it is possible that there are atypicalities specifically related to the integration of implicit (inferred) information into the discourse model. This process is essential for an inference to be formed and if atypical, could contribute to the reports of poorer performance for inferencing and comprehension tasks.

This experiment examined the on-line formation of co-referential links during reading in ASD. Recall that co-referential links often require readers to make a basic inference and difficulties associated with computation of

co-reference are likely to contribute to reading comprehension in ASD. The formation of anaphoric links is a common type of co-referential processing that occurs when a co-reference relation exists between a word and another word that appeared previously within the text. Previous studies have demonstrated that reading times and fixations upon anaphoric category nouns (e.g., *bird*) are longer, following an atypical exemplar (e.g., *penguin*), in comparison to a typical exemplar [e.g., *pigeon*; Duffy & Rayner, 1990; Garrod & Sanford, 1977; Myers, Cook, Kambe, Mason, & O'Brien, 2000; Rayner, Kambe, & Duffy, 2000; c.f. Van Gompel, Liversedge, & Pearson, 2004]. This increased fixation time is thought to reflect the greater difficulty associated with inferring and forming a link between nouns and atypical exemplars that are semantically less well connected than typical exemplars. By adopting this paradigm, we can examine the efficiency with which a very basic inference is computed (i.e., that *bird* co-refers to *pigeon/penguin*) and is used to form a co-referential link that is incorporated into the discourse model in ASD. If readers with ASD compute an anaphoric link less efficiently than TD readers, then we predict interactive effects, whereby TD readers demonstrate standard typicality effects upon fixation of the category noun, but these effects are reduced, less immediate, or even absent in individuals with ASD.

Method

Participants

Two groups of native English speakers were recruited who had no learning difficulties (e.g., dyslexia). One group consisted of 16 adults (one female) with a formal diagnosis of an ASD, 14 were diagnosed with Asperger's syndrome, one with autism and one with pervasive developmental disorder. These participants were recruited through local charitable organisations on a voluntary basis. ASD diagnoses were confirmed using module 4 of the ADOS-2 [Lord et al., 2012], administered by the first author who has received accreditation to administer this assessment for research purposes. The TD group consisted of 16 volunteers from the local community (one female) recruited through online advertisement. All participants were paid for their time. The participant groups did not differ in age t $(29.77) = 0.88$, $P = .387$, verbal IQ t $(28.98) = 0.16$, $P = .876$, performance IQ t $(26.00) = 0.17$, $P = .863$ or full scale IQ t $(27.51) = 0.44$, $P = .662$, as assessed using the Weschler abbreviated scale of intelligence[2]

---

[2]One ASD participant could not complete all 4 subtests because of time constraints, and therefore verbal and performance IQ could not be estimated. Full scale IQ was estimated based on performance on the vocabulary and matrix reasoning subtests.

Table I. Means (Standard Deviations) for ASD and TD Group's Age, Self-Reported Autistic Traits, Intelligence, Expressive Language and Reading Skill

Measure	TD	ASD
Age	33.69 (10.58)	37.13 (11.55)
Autism quotient	17.44 (9.54)	35.81 (8.18)
Full scale IQ	118.25 (10.04)	116.38 (13.17)
Performance IQ	117.06 (9.83)	116.33 (13.04)
Verbal IQ	116.00 (11.08)	116.60 (10.11)
Expressive language (max score 96)	91.19 (5.11)	85.19 (6.12)
Single word reading (max score 70)	68.25 (2.46)	68.13 (2.33)
Passage comprehension (max score 13)	10.06 (1.82)	8.38 (1.95)

Table II. Example of the Experimental Stimuli and Regions of Interest

The instruction booklet said that Tom would need a[1] hammer/plunger[2] to fix the kitchen sink.[3]
He borrowed the[4] tool[5] from his[6] next door neighbour.[7]

1 = start; 2 = antecedent; 3 = post-antecedent; 4 = pre-target; 5 = target; 6 = post target; 7 = end.

[Wechsler, 1999]. All participants completed the recalling subtest from the clinical evaluation of language fundamentals [CELF; Semel, Wiig, & Secord, 2003] that is sensitive to language impairment [e.g., Conti-Ramsden, Botting, & Faragher, 2001] and scored above the highest age equivalent cut off (>12.11 years). However, the ASD group's raw score (standard scores not available for this age range) was lower than the TD group's $t (29.08) = 3.01$, $P = .005$. The York Assessment of Reading Comprehension [Snowling et al. 2010] was administered to assess reading skill and the groups did not differ in accuracy scores for single word reading $t (29.91) = 0.15$, $P = .884$, but the ASD group had lower accuracy for a passage comprehension task, than TD controls $t (27.93) = 2.77$, $P = .010$. As expected, the ASD participants reported a higher number of autistic traits, in comparison to the TD group $t (29.32) = 5.85$, $P < .001$ [Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001]. See Table I for group means and standard deviations.

Materials

Forty mini discourses (two sentences) were developed that included an exemplar noun in the first sentence that co-referred to a category noun in the second sentence. The exemplar was either a typical or an atypical instance of the category. Stimuli were divided into seven regions, two of which were critical regions of interest. This included the target region that consisted of the category noun, and the post target region that consisted of one long, or two short words (see Table II).

Twenty category nouns and associated exemplars were chosen on the basis of Van Overschelde, Rawson, and Dunlosky's [2004] updated version of Battig and Montague's [1969] category noun typicality norms, and 20 were selected based upon the first author's judgement. To ensure all exemplars were correctly categorised as typical or atypical, 16 undergraduates listed as many instances of each category they could, in the order they thought of them [e.g., see Van Overschelde et al. 2004]. It was assumed that instances listed by a high percentage of participants were typical and those listed by few were atypical. From the original 40 stimuli, 32 stimuli pairs that differed in typicality were selected (proportion of participants that listed each instance type; typical $M = 0.84$, $SD = 0.15$, atypical $M = 0.04$, $SD = 0.03$, see Appendix for a full list of the experimental stimuli). The stimuli were split into two lists that each contained one version of each stimulus (16 typical, 16 atypical). Each participant viewed one list. Thus, sentences containing typical and atypical instances of the category were rotated across subject groups according to a Latin Square.

Apparatus

The text was displayed on a 21inch CRT monitor (100 Hz). Eye movements were monitored using a desktop mounted Eyelink 1000 (1,000 Hz, SR Research, Ottawa, Canada) and head movements were minimised using a chin and forehead rest. A 13-point calibration procedure was used, and fixations on each point were required to be within 0.5 degrees of error. Calibration accuracy was checked regularly, and recalibration was performed when needed.

Procedure

The text was presented across two lines and the target category noun was positioned at approximately the middle of the second line. In total participants read 67 discourses, 5 were for practice and occurred at the beginning of the session, 32 were experimental discourses, and 30 were filler discourses. The experimental and filler trials were presented in random order. Participants were asked to read normally for comprehension and to answer a *Yes/No* comprehension question (for which *Yes* and *No* answers were equally likely) after half of the trials using a button controller (e.g., *Did Tom own the tool?*).

Design

A 2×2 (Typicality: typical vs. atypical) (Group: TD vs. ASD) design was employed with typicality as a within participants factor, and group as a between subjects factor. Definitions of eye movement and other measures examined are included in the "Results" section.

Table III. Model Parameters, Means (Standard Deviations) for Global Analyses

Measure		Model *b*	SE	*t*	*P*	TD Typical	TD Atypical	ASD Typical	ASD Atypical
Accuracy[a] (proportion correct)	Intercept	**4.15**	**0.60**	**6.93**	**<.001**	0.97 (0.17)	0.96 (0.19)	0.98 (0.12)	0.99 (0.15)
	Group	0.85	0.73	1.16	.248				
	Typicality	0.52	0.58	−0.57	.571				
	Expressive	0.03	0.05	0.56	.576				
	Interaction	−0.19	1.16	−0.16	.872				
Mean fixation duration (msec)	Intercept	**204.06**	**3.35**	**60.83**	**<.001**	200 (23)	200 (22)	207 (26)	209 (25)
	Group	3.12	7.55	0.41	.682				
	Typicality	1.66	1.04	1.60	.112				
	Expressive	−0.78	0.60	−1.30	.205				
	Interaction	2.02	2.07	0.97	.331				
Fixation count (count)	Intercept	**25.53**	**1.05**	**24.27**	**<.001**	22 (7)	23 (6)	28 (9)	28 (8)
	Group	3.51	2.17	1.61	.117				
	Typicality	0.42	0.36	1.16	.251				
	Expressive	**−0.36**	**0.17**	**−2.18**	**.038**				
	Interaction	0.05	0.71	0.07	.944				
Total reading time (msec)	Intercept	**6,422.97**	**301.59**	**21.30**	**<.001**	5,368 (1,793)	5,510 (1,716)	7,225 (2,567)	7,289 (2,407)
	Group	1,257.26	631.28	1.99	.055				
	Typicality	119.50	94.89	1.26	.213				
	Expressive	**−107.70**	**47.57**	**−2.26**	**.031**				
	Interaction	−35.68	190.79	−0.19	.852				

Note. Expressive = expressive language score (CELF); Interaction = group X typicality interaction term.

[a] Values in the t column correspond to z values.

Results

Data Preparation and Analysis

Eye movement, accuracy, and reading time data were analysed with linear mixed effect models[3] [Baayen, Davidson, & Bates, 2008] using the lme4 library [Bates, Maechler, Bolker, & Walker, 2015] in R [version 3.2.4; R Core Team, 2016]. Group and typicality were coded as categorical fixed effects to attain main effects, specified using the contra.sdif function from the MASS library [Venables & Ripley, 2002]. In addition, expressive language scores were included as a continuous fixed effect (centred), given the differences detected between groups on this measure. This was to ensure that any differences as a result of expressive language were not misattributed to be a consequence of ASD. The full random structure was included [Barr, Levy, Scheepers, & Tily, 2013], with intercepts allowed to vary for each participant and stimuli number. Random slopes were included at the participant level for typicality, and at the item level for group, typicality, and expressive language, resulting in the following syntax: lmer (dv ~ group*typicality + centred_express + (1+ typicality |participant_ID) + (1+ group*typicality + centred_express |stimuli_no), data). Therefore, for each measure, one model was run. When

models did not converge, parameters were systematically removed from the random structure, beginning with the items level correlation. If a model still did not converge, this was re-entered and the model was re-run excluding the interaction. If this was still unsuccessful both the correlation and interaction were removed, and as necessary, each random slope was removed one-by-one in the following order; expressive language, typicality, group. If the model still did not converge, the same procedure was followed for the participant level of the random structure.

Accuracy and Reading Time Analyses

Recall that participants had to answer comprehension questions about the content of the discourses following 50% of trials. These questions were not related to the co-referential link. Comprehension accuracy was high and there was no effect of group, typicality or expressive language.

For overall reading time (time from sentence onset until participants pressed a button to indicate they had finished reading), there was no effect of typicality, but there was an effect of expressive language. Participants with lower expressive scores took longer to read the text. There was also a numerical tendency for ASD readers to take longer to read the text than TD controls. Although not statistically reliable, this trend is consistent with previous reports of ASD readers engaging in increased re-reading [e.g., Au-Yeung et al., 2015; Howard et al., 2017a; Sansosti et al., 2013] and is examined

Table IV. Model Parameters, Means (Standard Deviations) in msec for Target Region Analyses

Measure		Model				TD		ASD	
		b	SE	*t*	*P*	Typical	Atypical	Typical	Atypical
First fixation duration	Intercept	**199.49**	**3.80**	**52.44**	**<.001**	199 (54)	202 (57)	197 (57)	201 (62)
	Group	−8.44	9.33	−0.91	.371				
	Typicality	4.06	4.24	0.96	.342				
	Expressive	−1.11	0.72	−1.55	.131				
	Interaction	1.30	7.81	0.17	.868				
Single fixation duration	Intercept	**200.86**	**4.16**	**48.25**	**<.001**	199 (55)	204 (57)	197 (58)	200 (62)
	Group	−9.85	9.95	−0.99	.329				
	Typicality	5.77	4.63	1.24	.219				
	Expressive	−1.32	0.78	−1.69	.099				
	Interaction	−0.50	9.07	−0.06	.956				
Gaze duration	Intercept	**215.22**	**4.79**	**44.92**	**<.001**	210 (66)	224 (76)	211 (72)	218 (78)
	Group	−15.00	10.98	−1.37	.181				
	Typicality	**11.16**	**5.27**	**2.12**	**.037**				
	Expressive	**−2.00**	**0.82**	**−2.44**	**.021**				
	Interaction	−8.45	10.28	−0.82	.412				
Total time	Intercept	**254.61**	**9.05**	**28.12**	**<.001**	239 (94)	236 (86)	275 (129)	275 (134)
	Group	28.04	17.34	1.62	.116				
	Typicality	0.60	8.03	0.08	.941				
	Expressive	−1.33	1.34	−0.99	.330				
	Interaction	5.27	15.95	0.33	.743				

Note. Expressive = expressive language score (CELF); Interaction = group X typicality interaction term.

in more detail in the supplementary analyses. See Table III for means and model parameters.

Eye Tracking Data Analyses

Global analyses. To determine whether there were any basic sampling differences between the two groups, we firstly considered global measures of processing (see Table III for global analyses means and model parameters). For average fixation duration, there was no effect of group, typicality, or expressive language. Thus, on average, both groups extracted information at a similar speed and this was not influenced by verbal skill. For average number of fixations, there was no difference between groups or typicality conditions, but there was a reliable effect of expressive language. This was a negative association, where as expressive language score decreased, number of fixations increased. This indicates that both ASD and TD readers with reduced verbal language proficiency made more fixations in order to comprehend the texts.

Target word analyses. For first fixation durations (the duration of the first fixation upon the target word) and single fixation durations (the duration of a fixation when there is only one fixation made upon the target during first pass reading), there was no effect of group, typicality, or expressive language skill. For gaze duration (the sum of fixations upon the target word, from the first until the eyes leave this region to either the left or right), an effect of typicality was found, with

both groups having longer gaze durations upon the category noun when it's antecedent was atypical, in comparison to typical (11 msec effect). There was no difference between the groups, but there was an effect of expressive language, whereby participants with low expressive language scores had longer gaze durations. This is suggestive that those with poorer expressive language skill (both TD and ASD) took longer to compute anaphoric links. For total times (total amount of time spent fixating the target, including revisits to this region), there was no effect of group, typicality or expressive language score. See Table IV for target word analyses means and model parameters.

Bayes factor analyses. One disadvantage of null hypothesis statistical tests is that only evidence *against* the null hypothesis can be attained [Wagenmakers, 2007]. However, Bayesian hypothesis testing (Bayes Factor) allows one to comparatively assess the relative evidence for opposing models/hypotheses (including null effects). Bayes Factor is the ratio of the data's marginal likelihood under each hypothesis/model (e.g., $p(D|H_1)/p(D|H_0) = BF_{10}$, where D represents observed data, and the subscript denotes the hypotheses [*H*] being compared), and reflects the change of prior odds (belief about hypothesis before observing data) relative to posterior odds, based upon the observed data. A BF_{10} value larger than 1 would suggest the evidence is in favour of H_1, whereas a BF_{10} smaller than 1 would suggest evidence in favour of H_0 [for an in depth discussion of

Table V. Model Parameters, Means (Standard Deviations) in msec for Post Target Region Analyses

Measure		Model				TD		ASD	
		b	SE	*t*	*P*	Typical	Atypical	Typical	Atypical
First fixation duration	Intercept	**205.03**	**4.24**	**48.36**	**<.001**	200 (54)	207 (57)	207 (59)	208 (58)
	Group	−1.54	9.03	−0.17	.866				
	Typicality	4.75	3.72	1.28	.210				
	Expressive	−0.88	0.72	−1.23	.229				
	Interaction	−5.29	7.12	−0.74	.458				
Single fixation duration	Intercept	**214.07**	**5.85**	**36.57**	**<.001**	207 (57)	213 (63)	212 (63)	207 (56)
	Group	−7.88	12.02	−0.66	.517				
	Typicality	0.94	5.72	0.16	.872				
	Expressive	−1.14	0.92	−1.24	.224				
	Interaction	−8.34	10.26	−0.81	.426				
Gaze duration	Intercept	**284.99**	**13.80**	**20.65**	**<.001**	290 (137)	279 (119)	293 (138)	289 (131)
	Group	5.12	28.19	0.18	.857				
	Typicality	−5.58	10.51	−0.53	.600				
	Expressive	−0.09	2.25	−0.04	.969				
	Interaction	5.09	19.32	0.26	.794				
Total time	Intercept	**366.01**	**19.33**	**18.94**	**<.001**	334 (175)	334 (159)	403 (231)	399 (216)
	Group	58.55	36.99	1.58	.124				
	Typicality	−2.40	12.23	−0.20	.845				
	Expressive	−1.34	2.86	−0.47	.641				
	Interaction	−6.19	23.82	−0.26	.795				

Note. Expressive = expressive language score (CELF); Interaction = group X typicality interaction term.

Bayes Factor, interested readers are referred to Kass & Raftery, 1995; Rouder, Morey, Speckman, & Province, 2012; Wagenmakers, 2007].

Given the small sample size and the effect of typicality for gaze durations upon the target word appearing to be numerically smaller for the ASD group, in comparison to the TD group, we carried out Bayes Factor analyses [Kass & Raftery, 1995; Rouder et al., 2012]. This was to examine whether the lack of difference between groups was a "true" null effect, or a Type II error. Therefore, we compared the relative evidence for the original model that included a group by typicality interaction, to a model that did not include group as a fixed effect (null effect of group—only typicality and expressive language were included as fixed effects). This was calculated using the BayesFactor Package [Morey & Rouder, 2015] in R [R Core Team, 2016], with 100,000 Monte Carlo iterations and *g*-priors scaled to *r* = .5 for fixed effects, as is recommended for small effect sizes [Rouder et al., 2012]. BayesFactor was < 0.001, which indicates strong evidence in favour of the model excluding group as a predictor (null effect of group), according to Jeffreys [1961] evidence categories. Based upon this analysis, we can conclude that TD and ASD readers did not differ in the immediacy with which they computed an anaphoric link.

Post target analyses. Previous studies examining referential processing have occasionally found effects to occur or to continue to occur on words following the target. However, we found no reliable effects for this region. Fixed effect model parameters along with group means and standard deviations are presented in Table V.

Supplementary analyses. To examine the time course of the numerically increased reading times for ASD readers, a series of supplementary analyses was run. This aimed to identify whether the increased reading time occurred during first-pass (fixations made on a region, or regressions initiated out of a region, prior to the eyes leaving that region to the right to process "new" information), or at a later stage of processing (e.g., re-reading that occurs after the entire sentence has been read). Increased first-pass times would be indicative of a difficulty constructing an initial representation of the text, whereas increased second-pass times would be indicative of ASD readers taking longer to evaluate their interpretation of the text. For conciseness, we only report reliable group differences and interactions below. Means, standard deviations, and full model parameters are reported in Tables VI–VIII.

To examine first-pass times, gaze durations and the proportion of first-pass regressions (prior to fixating a later region) made out of each region was calculated. Recall that there were no differences between groups for target or post target gaze durations. Similarly, there were no differences in gaze duration between groups for any other region. For first pass regressions, there was a marginal effect of group in the post target region, with a trend suggesting ASD readers regressed more

Table VI. Model Parameters for Supplementary Analyses of the Regions Within the First Sentence

		Start				Antecedent				Post antecedent			
		b	SE	*t/z*	*P*	*b*	SE	*t/z*	*P*	*b*	SE	*t/z*	*P*
GD	Intercept	**1,081.33**	**90.01**	**12.01**	**<.001**	**289.19**	**13.38**	**21.62**	**<.001**	**800.23**	**67.19**	**11.91**	**<.001**
	Group	1.69	132.64	0.01	.990	−5.73	23.98	−0.24	.812	8.55	88.05	0.10	.923
	Typicality	−23.73	25.55	−0.93	.362	**93.10**	**19.25**	**4.84**	**<.001**	−51.92	33.26	−1.56	.128
	Expressive language	−12.89	9.88	−1.31	.202	−1.83	1.71	−1.07	.291	−9.60	6.83	−1.41	.170
	Typicality X group	−41.30	48.57	−0.85	.403	20.25	28.71	0.71	.485	−94.81	70.24	−1.35	.186
RO	Intercept	–	–	–	–	**−1.76**	**0.20**	**−8.93**	**<.001**	**−1.51**	**0.16**	**−9.74**	**<.001**
	Group	–	–	–	–	0.63	0.39	1.61	.107	0.44	0.27	1.59	.112
	Typicality	–	–	–	–	−0.18	0.31	−0.58	.560	**0.49**	**0.22**	**2.25**	**.025**
	Expressive language	–	–	–	–	0.00	0.03	−0.11	.914	−0.01	0.02	−0.72	.471
	Typicality X group	–	–	–	–	−0.59	0.61	−0.97	.332	−0.02	0.40	−0.04	.965
PR	Intercept	−0.02	0.22	−0.08	.933	**−0.80**	**0.19**	**−4.09**	**<.001**	**−0.73**	**0.19**	**−3.76**	**<.001**
	Group	**0.95**	**0.48**	**2.00**	**.046**	**0.94**	**0.37**	**2.56**	**.011**	0.54	0.40	1.35	.177
	Typicality	0.11	0.16	0.68	.497	**0.82**	**0.23**	**3.58**	**<.001**	0.06	0.18	0.34	.737
	Expressive language	**−0.09**	**0.04**	**−2.33**	**.020**	**−0.07**	**0.03**	**−2.36**	**.018**	**−0.07**	**0.03**	**−2.07**	**.038**
	Typicality X group	0.41	0.33	1.23	.220	0.16	0.41	0.39	.700	0.11	0.34	0.33	.744
RT	Intercept	**5.87**	**0.08**	**74.09**	**<.001**	**324.26**	**18.18**	**17.84**	**<.001**	**625.76**	**58.98**	**10.61**	**<.001**
	Group	**0.41**	**0.17**	**2.43**	**.022**	44.63	37.38	1.19	.244	122.80	91.58	1.34	.191
	Typicality	−0.08	0.08	−1.09	.289	61.75	33.38	1.85	.071	94.12	55.18	1.71	.091
	Expressive language	−0.02	0.01	−1.40	.173	−4.32	3.20	−1.35	.189	−2.61	6.60	−0.40	.696
	Typicality X group	−0.10	0.16	−0.63	.536	70.19	59.93	1.17	.245	105.12	120.9	0.87	.390

Note. GD = gaze duration; PR = proportion of re-reading; RO = proportion of first pass regressions out; RT = re-reading time.

Table VII. Model Parameters for Supplementary Analyses of the Regions Within the First Sentence

		Pre-target				Target				Post target				End			
		b	SE	*t/z*	*P*	*b*	SE	*t/z*	*P*	*b*	SE	*t/z*	*P*	*b*	SE	*t/z*	*P*
GD	Intercept	**713.84**	**56.51**	**12.63**	**<.001**	–	–	–	–	–	–	–	–	**615.37**	**44.23**	**13.91**	**<.001**
	Group	19.39	88.80	0.22	.829	–	–	–	–	–	–	–	–	61.18	56.08	1.09	.284
	Typicality	29.04	21.80	1.33	.192	–	–	–	–	–	–	–	–	−41.44	22.13	−1.87	.069
	Expressive	−4.87	6.94	−0.70	.489	–	–	–	–	–	–	–	–	−5.89	4.52	−1.30	.203
	Interaction	−3.16	36.07	−0.09	.931	–	–	–	–	–	–	–	–	−9.58	42.01	−0.23	.820
RO	Intercept	**−4.93**	**0.64**	**−7.73**	**<.001**	**−1.57**	**0.20**	**−8.04**	**<.001**	**−2.83**	**0.28**	**10.08**	**<.001**	**0.58**	**0.22**	**2.58**	**<.001**
	Group	1.07	0.62	1.73	.083	0.28	0.39	0.71	.481	1.05	0.55	1.91	.056	0.18	0.49	0.37	.715
	Typicality	**1.75**	**0.50**	**3.46**	**.001**	−0.35	0.21	−1.67	.095	0.43	0.51	0.83	.406	−0.10	0.16	−0.66	.512
	Expressive	−0.06	0.04	−1.53	.125	−0.01	0.03	−0.44	.662	0.00	0.04	0.12	.904	**−0.12**	**0.04**	**−2.98**	**.003**
	Interaction	−1.12	0.95	−1.17	.242	0.03	0.38	0.07	.943	−0.50	0.92	−0.54	.591	0.41	0.32	1.30	.193
PR	Intercept	**−1.03**	**0.23**	**−4.48**	**<.001**	**−1.90**	**0.23**	**−8.22**	**<.001**	**−1.05**	**0.17**	**−6.28**	**<.001**	**−1.43**	**0.26**	**−5.52**	**<.001**
	Group	**1.06**	**0.44**	**2.41**	**.016**	**1.35**	**0.42**	**3.26**	**.001**	0.51	0.36	1.44	.150	0.85	0.54	1.58	.114
	Typicality	−0.42	0.27	−1.54	.123	**−0.68**	**0.30**	**−2.27**	**.023**	0.06	0.19	0.34	.731	−0.05	0.25	−0.22	.827
	Expressive	−0.04	0.03	−1.34	.180	0.04	0.03	1.16	.245	0.00	0.03	0.14	.887	−0.03	0.04	−0.82	.415
	Interaction	0.50	0.38	1.30	.193	0.48	0.54	0.90	.369	−0.19	0.34	−0.56	.573	0.17	0.46	0.36	.717
RT	Intercept	**417.58**	**47.74**	**8.75**	**<.001**	**231.80**	**14.49**	**16.00**	**<.001**	**303.16**	**11.93**	**25.41**	**<.001**	**632.53**	**48.59**	**13.02**	**<.001**
	Group	172.14	86.48	1.99	.056	15.51	28.58	0.54	.595	**57.97**	**26.78**	**2.17**	**.045**	89.76	95.48	0.94	.358
	Typicality	39.37	68.03	0.58	.567	20.47	25.25	0.81	.423	−9.71	22.87	−0.43	.672	−39.57	62.69	−0.63	.530
	Expressive	−4.76	5.80	−0.82	.421	−1.66	2.15	−0.77	.451	−3.20	2.03	−1.58	.135	−1.80	6.73	−0.27	.792
	Interaction	16.70	114.92	0.15	.885	34.17	44.07	0.78	.442	−14.90	44.62	−0.33	.739	177.49	122.62	1.45	.153

Note. GD = gaze duration; PR = proportion of re-reading; RO = proportion of first pass regressions out; RT = re-reading time.

frequently than TD readers, however, there was no group difference for any other region. This suggests that first pass processing of the text was very similar between participant groups.

To examine the second possibility that the increased reading times reflected re-reading that occurred after first pass, re-reading time was calculated for each region (i.e., total time minus gaze duration). This measure contained a large proportion of zero values (when no re-reading occurred) and therefore, we considered these data in two different ways. First, we computed a binomial variable representing the proportion of re-reading

Table VIII. Means (Standard Deviations) for All Measures Calculated for the Supplementary Analyses

			GD	RO	PR	RT
Start	TD	Typical	1,025 (534)	–	0.36 (0.48)	381 (284)
		Atypical	1,018 (510)	–	0.35 (0.48)	363 (270)
	ASD	Typical	1,133 (685)	–	0.60 (0.49)	823 (711)
		Atypical	1,084 (617)	–	0.65 (0.48)	631 (486)
Antecedent	TD	Typical	245 (85)	0.15 (0.35)	0.19 (0.39)	291 (144)
		Atypical	330 (165)	0.16 (0.37)	0.28 (0.45)	310 (230)
	ASD	Typical	245 (104)	0.26 (0.44)	0.39 (0.49)	313 (181)
		Atypical	356 (215)	0.21 (0.41)	0.56 (0.50)	415 (301)
Post antecedent	TD	Typical	745 (479)	0.14 (0.35)	0.27 (0.44)	460 (589)
		Atypical	746 (484)	0.20 (0.40)	0.27 (0.45)	498 (602)
	ASD	Typical	857 (612)	0.21 (0.41)	0.44 (0.50)	525 (691)
		Atypical	757 (528)	0.29 (0.46)	0.46 (0.50)	647 (827)
Pre-target	TD	Typical	664 (316)	0.01 (0.09)_	0.24 (0.43)	307 (197)
		Atypical	704 (375)	0.05 (0.23)	0.20 (0.40)	382 (377)
	ASD	Typical	722 (408)	0.04 (0.20)	0.45 (0.50)	509 (397)
		Atypical	749 (411)	0.10 (0.30)	0.43 (0.50)	681 (584)
Target	TD	Typical	–	0.20 (0.40)	0.14 (0.34)	217 (75)
		Atypical	–	0.15 (0.36)	0.09 (0.28)	231 (100)
	ASD	Typical	–	0.28 (0.45)	0.27 (0.44)	226 (87)
		Atypical	–	0.22 (0.41)	0.24 (0.43)	276 (171)
Post target	TD	Typical	–	0.05 (0.21)	0.22 (0.41)	268 (159)
		Atypical	–	0.07 (0.26)	0.26 (0.44)	266 (117)
	ASD	Typical	–	0.12 (0.33)	0.33 (0.47)	352 (198)
		Atypical	–	0.14 (0.34)	0.34 (0.47)	331 (184)
End	TD	Typical	581 (365)	0.56 (0.50)	0.18 (0.39)	684 (411)
		Atypical	545 (311)	0.50 (0.50)	0.21 (0.41)	536 (354)
	ASD	Typical	690 (412)	0.67 (0.47)	0.33 (0.47)	730 (526)
		Atypical	634 (434	0.68 (0.47)	0.35 (0.48)	745 (496)

Note. GD = gaze duration in msec; PR = proportion of re-reading; RO = proportion of first pass regressions out; RT = re-reading time in msec.

for each region. Second, we calculated how long participants spent re-reading each region, when they re-read (i.e., zero re-reading times were removed). Participants with ASD re-read the start, antecedent, pre-target, and target regions on a higher proportion of trials, in comparison to TD readers. In addition, ASD participants had longer re-reading times for the start and post target regions, and marginally longer re-reading times for the pre-target region, in comparison to TD readers. These analyses indicate that whilst ASD and TD readers' first-pass reading was very comparable, ASD participants re-read the text more frequently and spent longer doing so, than TD readers.

Discussion

In this experiment, we examined the efficiency of co-reference computation during reading in ASD. Participants read mini discourses containing an anaphoric link between category nouns and typical or atypical exemplars. No group differences or interactions were detected. Both groups had longer gaze durations at the category noun when it was preceded by an atypical exemplar, in comparison to a typical exemplar. This is consistent with previous studies that have manipulated typicality [e.g., Duffy & Rayner, 1990; Myers et al., 2000; Rayner, et al., 2000] and suggests that the efficiency with which ASD readers computed a very basic inference and formed a co-reference link between instance and category words, was comparable to TD readers. Our results do not support the hypotheses that the integration of information between sentences or the incorporation of an implicit link into the discourse model is impaired in ASD (as per CIP, WCC theories). It therefore seems that comprehension difficulties reported in the literature do not arise due to ASD readers less immediately establishing co-referential links. Instead, our findings are consistent with recent studies reporting intact integrative processing during reading in ASD [e.g., Au-Yeung et al., 2015; Howard et al., 2017a].

These results are in contrast to several of the studies considered in the introduction, that report reduced performance accuracy and processing efficiency in ASD for tasks that require an inference to be formed. It is possible that such results occur for studies where participants are required to form causal and pragmatic inferences [e.g., Bodner et al., 2015; Dennis et al., 2001; Jolliffe & Baron-Cohen, 1999, 2000; Micai et al., 2017; Minshew et al., 1995; Norbury & Bishop, 2002; Norbury & Nation, 2011; Tirado & Saldaña, 2016]. As

noted earlier, forming such inferences could be considered more cognitively demanding relative to the type of inference that participants computed in the current study. In the present study, participants needed only to form a referential link between two words on the basis of lexical/semantic knowledge (e.g., that a *penguin/pigeon* is a *bird*). In contrast, causal inferences require a reader to infer an event structure that captures a causality relation, usually derived from situational world knowledge. One explanation for the inconsistency between our own and previous work is that inferential processing is not universally atypical in ASD, but may vary dependent upon the degree of processing complexity associated with the formation of a particular inference. Moreover, differences in the efficiency of the use of situational world knowledge have been recently reported during on-line reading in ASD [Howard, Liversedge, & Benson, 2017b] and the requirement to engage with situational world knowledge may therefore be a factor that modulates inferential processing in ASD.

Potential differences in world knowledge processing, however, cannot account for why our findings contrast with O'Connor and Klein's [2004] report that comprehension accuracy scores for adolescents with ASD improve when prompts to make anaphoric links are provided. This finding implies that the spontaneous formation of anaphoric links is less efficient in ASD, and is inconsistent with the present findings. There are two possible reasons why our data may not coincide. Firstly, we tested an adult sample, whereas O'Connor and Klein [2004] recruited adolescents. Perhaps the development of co-referential processing is delayed in ASD, and this difference diminishes with age. Consistent with this suggestion, Bodner et al. [2015] reported the difference in inferencing accuracy between TD and ASD participants to reduce as age increases. Alternatively, it is possible that differences in verbal language proficiency could account for O'Connor and Klein's [2004] findings. O'Connor & Klein [2004] assessed verbal language proficiency using the Test of Language Development and these scores were found to correlate negatively with comprehension scores during the anaphoric cueing task. Therefore, perhaps this intervention was more beneficial to students with lower verbal language proficiency, and that verbal proficiency underpinned anaphoric processing differences in O'Connor and Klein's [2004] sample, as opposed to ASD per se.

The measure of expressive language we used (sentence repetition; CELF) taps into multiple processes related to verbal language proficiency, such as working memory, phonological and syntactic processing. Lower scores on this measure were predictive of participants making a higher number of fixations and having longer overall reading times. These data suggest that reduced verbal language proficiency was related to general reading proficiency, and this is consistent with previous studies that have found expressive language to be predictive of reading skill in both TD and ASD participants [e.g., Lucas & Norbury, 2014; Norbury & Nation, 2011]. In addition, expressive language was predictive of gaze duration on the target words, which is where we first observed evidence of the formation of the anaphoric link in the eye movement record. This suggests that verbal language skill is also associated with inferential processing, and again is consistent with previous findings [e.g., Lucas & Norbury, 2015; Pérez, Joseph, Bajo, & Nation, 2016; Singer, Andruslak, Reisdorf, & Black, 1992; Singer & Ritchot, 1996].

Consistent with previous studies that have examined eye movements during reading in ASD, we found ASD participants to be more likely to re-read than TD participants [Au-Yeung et al., 2015; Howard et al., 2017a,b]. The likelihood that ASD and TD readers would initiate a first-pass regression was comparable, however the likelihood of re-reading occurring following first-pass, was greater in the ASD readers, than the TD readers. Re-reading was not concentrated on any particular region, rather, increased re-reading was found for the majority of regions. Moreover, ASD participants were as likely to re-read in both typical and atypical conditions. Therefore, this re-reading behaviour is unlikely to reflect ASD participants having difficulty engaging in basic linguistic processing required for the initial construction of a mental representation of the text [e.g., Howard et al., 2017a,b]. Instead, it seems likely that this re-reading may be a result of increased evaluation of the text content. We have speculated in previous papers that this may be a "cautious" or "checking" strategy that readers with ASD adopt. However, it is also quite possible that this is a task effect [see Howard et al., 2017a], or alternatively, that re-reading reflects a repetitive behaviour that is characteristic of ASD. These possibilities remain to be empirically investigated. Suffice to say, the current findings represent another independent demonstration of increased re-reading in adults with ASD.

In summary, we have demonstrated that typicality effects associated with the formation of a co-reference link between a category noun and an instance of that category occur with the same immediacy in ASD readers as in TD readers. This result is inconsistent with the suggestion that integrative processes required for the computation of co-reference are impaired in ASD readers.

Ethical Statement

The University of Southampton's Ethics and Research Governance Board ethically approved this research and

all participants gave informed written consent before taking part. The authors declare that they have no conflict of interest.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Åsberg, J., Kopp, S., Berg-Kelly, K., & Gillberg, C. (2010). Reading comprehension, word decoding and spelling in girls with autism spectrum disorders (ASD) or attention-deficit/hyperactivity disorder (AD/HD): Performance and predictors. International Journal of Language & Communication Disorders, 45, 61–71.

Au-Yeung, S.K., Kaakinen, J.K., Liversedge, S.P., & Benson, V. (2015). Processing of written irony in autism spectrum disorder: An eye movement study. Autism Research, 8, 749–760. doi:10.1002/aur.1490

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59, 390–412.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31, 5–17.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67, 1–48.

Battig, W.F., & Montague, W.E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80, 1.

Bodner, K.E., Engelhardt, C.R., Minshew, N.J., & Williams, D.L. (2015). Making inferences: Comprehension of physical causality, intentionality, and emotions in discourse by high-functioning older children, adolescents, and adults with autism. Journal of Autism and Developmental Disorders, 45, 2721–2733.

Brown, H.M., Oram-Cardy, J., & Johnson, A. (2013). A meta-analysis of the reading comprehension skills of individuals on the autism spectrum. Journal of Autism and Developmental Disorders, 43, 932–955.

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

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

Duffy, S.A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent typicality and distance. Language and Speech, 33, 103–119.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. Journal of Verbal Learning and Verbal Behavior, 22, 75–87.

Frith, U., & Happé, F. (1994). Autism - Beyond the theory of mind. Cognition, 50, 115–132.

Garrod, S., & Sanford, A. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior, 16, 77–90.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. Psychological Review, 101, 371–395.

Haviland, S.E., & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 13, 512–521.

Howard, P.L., Liversedge, S.P., & Benson, V. (2017a). Benchmark eye movement effects during natural reading in autism spectrum disorder. Journal of Experimental Psychology: Learning, Memory & Cognition, 43, 109–127. doi: 10.1037/xlm0000289

Howard, P.L., Liversedge, S.P., & Benson, V. (2017b). Investigating the use of world knowledge during on-line comprehension in adults with autism spectrum disorder. Journal of Autism and Developmental Disorders, 47, 2039–2053.

Huemer, S.V., & Mann, V. (2010). A comprehensive profile of decoding and comprehension in autism spectrum disorders. Journal of Autism and Developmental Disorders, 40, 485–493. doi:10.1007/s10803-009-0892-3

Jeffreys, H. (1961). Theory of probability (3rd ed.). New York: Oxford University Press.

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: Linguistic processing in high-functioning adults with autism or Asperger syndrome: Is local coherence impaired? Cognition, 71, 149–185.

Jolliffe, T., & Baron-Cohen, S. (2000). Linguistic processing in high-functioning adults with autism or Asperger's syndrome. Is global coherence impaired? Psychological Medicine, 30, 1169–1187.

Jones, C.R.G., Happé, F., Golden, H., Marsden, A.J.S., Tregay, J., Simonoff, E., . . . Charman, T. (2009). Reading and arithmetic in adolescents with autism spectrum disorders: Peaks and dips in attainment. Neuropsychology, 23, 718–728.

Kass, R.E., & Raftery, A.E. (1995). Bayes factors. Journal of the American Statistical Association, 90, 773–795.

Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule (2nd ed.). Torrance, CA: Western Psychological Services.

Lucas, R., & Norbury, C.F. (2014). Levels of text comprehension in children with autism spectrum disorders (ASD): The influence of language phenotype. Journal of Autism and Developmental Disorders, 44, 2756–2768. doi:10.1007/s10803-014-2133-7

Lucas, R., & Norbury, C.F. (2015). Making inferences from text: It's vocabulary that matters. Journal of Speech, Language, and Hearing Research, 58, 1224–1232.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. Psychological Review, 99, 440–466.

Micai, M., Joseph, H., Vulchanova, M., & Saldaña, D. (2017). Strategies of readers with autism when responding to inferential questions: An eye-movement study. Autism Research, 10, 888–900.

Minshew, N.J., & Goldstein, G. (1998). Autism as a disorder of complex information processing. Mental Retardation and Developmental Disabilities Research Reviews, 4, 129–136.

Minshew, N.J., Goldstein, G., & Siegel, D.J. (1995). Speech and language in high-functioning autistic individuals. Neuropsychology, 9, 255–261.

Morey, R.D., & Rouder, J.N. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-2. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Myers, J.L., Cook, A.E., Kambe, G., Mason, R.A., & O'Brien, E.J. (2000). Semantic and episodic effects on bridging inferences. Discourse Processes, 29, 179–199.

Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. Journal of Autism and Developmental Disorders, 36, 911–919.

Norbury, C., & Nation, K. (2011). Understanding variability in reading comprehension in adolescents with autism spectrum disorders: Interactions with language status and decoding skill. Scientific Studies of Reading, 15, 191–210.

Norbury, C.F., & Bishop, D.V. (2002). Inferential processing and story recall in children with communication problems: A comparison of specific language impairment, pragmatic language impairment and high-functioning autism. International Journal of Language & Communication Disorders, 37, 227–251.

O'Connor, I.M., & Klein, P.D. (2004). Exploration of strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. Journal of Autism and Developmental Disorders, 34, 115–127.

Pérez, A., Joseph, H.S., Bajo, T., & Nation, K. (2016). Evaluation and revision of inferential comprehension in narrative texts: An eye movement study. Language, Cognition and Neuroscience, 31, 549–566.

Poynor, D.V., & Morris, R.K. (2003). Inferred goals in narratives: Evidence from self-paced reading, recall, and eye movements. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 3.

R Core Team. (2016). R: Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K., Kambe, G., & Duffy, S.A. (2000). The effect of clause wrap-up on eye movements during reading. The Quarterly Journal of Experimental Psychology: Section A, 53, 1061–1080.

Rouder, J.N., Morey, R.D., Speckman, P.L., & Province, J.M. (2012). Default Bayes factors for ANOVA designs. Journal of Mathematical Psychology, 56, 365–374. doi:10.1016/j.jmp.2012.08.001

Saldaña, D., Carreiras, M., & Frith, U. (2009). Orthographic and phonological pathways in hyperlexic readers with autism spectrum disorders. Developmental Neuropsychology, 34, 240–253.

Saldaña, D., & Frith, U. (2007). Do readers with autism make bridging inferences from world knowledge? Journal of Experimental Child Psychology, 96, 310–319.

Sansosti, F.J., Was, C., Rawson, K.A., & Remaklus, B.L. (2013). Eye movements during processing of text requiring bridging inferences in adolescents with higher functioning autism spectrum disorders: A preliminary investigation. Research in Autism Spectrum Disorders, 7, 1535–1542.

Semel, E., Wiig, E.H., & Secord, W.A. (2003). Clinical evaluation of language fundamentals (CELF-4) (4th ed.). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Singer, M., Andruslak, P., Reisdorf, P., & Black, N.L. (1992). Individual differences in bridging inference processes. Memory & Cognition, 20, 539–548.

Singer, M., & Ritchot, K.F. (1996). The role of working memory capacity and knowledge access in text inference processing. Memory & Cognition, 24, 733–743.

Snowling, M.J., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., … Hulme, C. (2010). York assessment of reading for comprehension: Passage reading - secondary version. London: GL assessment.

Tirado, M.J., & Saldaña, D. (2016). Readers with autism can produce inferences, but they cannot answer inferential questions. Journal of Autism and Developmental Disorders, 46, 1025–1037.

Van Gompel, R.P.G., Liversedge, S.P., & Pearson, J. (2004). Antecedent typicality effects in the processing of noun phrase anaphora. In M. Carreiras & C. Clifton, Jr. (Eds.), The online study of sentence comprehension: Eye tracking, ERP, and beyond (pp. 119–137). Hove, UK: Psychology Press.

Van Overschelde, J.P., Rawson, K.A., & Dunlosky, J. (2004). Category norms: An update and expanded version of the Battig and Montague (1969) norms. Journal of Memory and Language, 50, 289–335.

Venables, W.N., & Ripley, B.D. (2002). Modern applied statistics with S (4th ed.). New York, NY: Springer.

Wagenmakers, E. (2007). A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review, 14, 779–804.

Wechsler, D. (1999). Wechsler abbreviated scale of intelligence. New York, NY: The Psychological Corporation, Harcourt Brace & Company.

Appendix

Experimental Stimuli

The second sentence contains the target category noun (in bold). The first sentence contains the instance category manipulation (in italics). The first instance listed (e.g., pigeon) is a typical exemplar; the second instance listed (e.g., penguin) is an atypical exemplar.

1. Whilst relaxing in the garden on Saturday, Sally picked a *daisy/geranium* from the ground. She tucked the **flower** behind her dog's ear and took a photograph.
2. The family sat watching a *pigeon/penguin* that was shuffling around next to the water. The youngest child pointed at the **bird** and laughed.
3. Once a fortnight Jackie played the *guitar/glockenspiel* in a local jazz band. She carefully polished and tuned the **instrument** before each performance.
4. Charles met Carly's *sister/granddaughter* at a house warming party. He was surprised to hear the **relative** was planning to move to Spain.

5. Jane sighed as she unpicked the stiches from the *cotton/taffeta* item she was making. She had been warned the **fabric** would be difficult to work with.

6. Fred wanted to paint his bedroom walls bright *green/cyan* to match his favourite t-shirt. His mother said the **colour** was too garish.

7. Derek bid for the antique *table/futon* with excitement. There were a lot of dealers who wanted the **furniture** for their collections.

8. Ellen sliced the *apple/cantaloupe* and placed it in her lunch box. As a child she was allergic to the **fruit** but had thankfully had grown out of it.

9. For safety reasons, the *gun/Taser* was kept in a locked glass cabinet. The owners were afraid somebody would steal the **weapon** for criminal activity.

10. The young energetic *Labrador/Schnauzer* bounded into the kitchen covered in mud. The children weren't meant to let the **dog** outside when it was raining.

11. The instruction booklet said that Tom would need a *hammer/plunger* to fix the kitchen sink. He borrowed the **tool** from his next door neighbour.

12. Jeremy was a trainee *doctor/masseuse* at the local clinic. At times the **profession** could be very challenging.

13. Jake had recently begun playing *rugby/curling* for the University team. He was amazed by the strong following the **sport** had from members of the public.

14. Robin wanted to study *chemistry/genetics* at Cambridge University. He had always been fascinated by the **science** since he was a young boy.

15. Brett had bought his younger brother a *teddy/yoyo* for his birthday. He wrapped the **toy** up in yellow shiny paper before giving it to him.

16. The gardener planted the *peas/radish* much earlier in the year than usual. He was confident the **vegetable** was hardy enough to cope with the cold.

17. At school, the children inspected an *ant/earwig* through a microscope. The exact species of the **insect** was difficult to identify because of its odd colouring.

18. Sarah squealed with excitement as she held the *cat/snail* for the first time. Her mother had bought her the **pet** as a Christmas present.

19. The chef picked some fresh *thyme/sorrel* from his restaurants allotment. He was planning to use the **herb** to flavour some roasted vegetables.

20. The smell of *paprika/cloves* floated through the café and reminded Ralph of his travels in Asia. He used to buy the **spice** by the pound from the market.

21. Vivian was surprised at how much she enjoyed *geography/citizenship* at college. She had dreaded the **subject** when it first appeared on her timetable.

22. Transporting the old *car/tank* to the museum was difficult. On the small, winding country roads the **vehicle** held up a lot of traffic.

23. Whilst staying in Australia Linda had a *snake/terrapin* living near the pond in her garden. She would sit and watch the **reptile** whilst drinking her morning coffee.

24. The group of tourists visited the secluded *house/igloo* on Tuesday. They were told the **dwelling** had been abandoned for years.

25. Zac ordered *chicken/lobster* for his main course at the fancy restaurant. He thought that the **meat** had a unique taste and texture.

26. As soon as he got home Ben threw his *trousers/overalls* in the washing basket. He had noticed the **garment** was covered in oil.

27. Tessa selected the *ruby/topaz* for her hand designed pendant. She knew the **jewel** would sparkle brightly when she wore it.

28. The deli owner gave free tasters of *cheddar/roquefort* to his customers. Many said the **cheese** was the best they had ever tasted.

29. Freya eventually found her red *boot/clog* under the bed. She had no idea how the **shoe** had got there, it had been missing for months.

30. Neil had been a devoted *Christian/Quaker* since he was a child. He had grown up learning about the **religion** and enjoyed being part of the community.

31. Everyone stared in fear as Wayne struggled to get the *canoe/catamaran* across the lake. When the wind changed direction the **boat** almost capsized.

32. James had been addicted to *heroin/valium* for 2 years. The price of the **drug** had doubled in recent months.