RESEARCH ARTICLE

Eye-Tracking Reveals Absent Repetition Learning Across the Autism Spectrum: Evidence From a Passive Viewing Task

Sebastian B. Gaigg [ID], Marie K. Krug [ID], Marjorie Solomon [ID], Amanda Roestorf [ID], Claire Derwent [ID], Sophie Anns [ID], Dermot M. Bowler [ID], Susan Rivera, Christine Wu Nordahl, and Emily J. H. Jones [ID]

In the domain of memory, autism is characterized by difficulties in explicitly remembering the specific order of stimuli, whereas implicit serial order memory appears to be preserved. This pattern is of considerable interest because serial order memory is known to play a critical role in children's language development. Currently, however, few paradigms exist that can effectively probe serial order memory across heterogeneous groups of children, including those who are minimally verbal. We present two experiments, involving 39 adults (20 ASD; 19 TD) and 130 children (86 ASD; 44 TD), that address this issue using an eye-tracking paradigm, which simply required participants to "watch out for a bunny" that appeared in repeating sequences of screen locations. The adults in Experiment 1 all had normative IQs, whereas Experiment 2 included children with and without substantial language and intellectual difficulties. In both experiments gaze latencies and anticipatory fixations to the bunny indicated reliable repetition learning effects in the TD but not the ASD groups. Importantly, we were able to acquire reliable data from around half of the children with significant language impairments in Experiment 2, indicating that the paradigm can shed light on important learning processes in this under-represented group. We discuss the implications of these findings for theories of memory in ASD as well as for the utility of eye-tracking technology to probe repetition learning effects in autism. *Autism Res* 2020, 00: 1–18. © 2020 The Authors. *Autism Research* published by International Society for Autism Research and Wiley Periodicals LLC.

Lay Summary: Remembering the specific order of stimuli plays an important role in language development and is thought to be a source of difficulty for autistic individuals. Research in this area, however, rarely includes autistic participants who are minimally verbal. Here we develop an eye-tracking paradigm that demonstrates serial order learning difficulties across the autism spectrum. We discuss the implications of these findings for our understanding of the role of memory difficulties in the varied language profiles across the autism spectrum.

Keywords: eye movement; language impairment; learning; learning disabilities; memory

Introduction

Serial order memory, broadly defined, refers to our ability to learn and later retrieve the specific order of stimuli or events. In explicit memory, this might be demonstrated by our ability to remember the order of digits of someone's phone number, while in implicit memory this might be indicated by our increasing proficiency at carrying out sequences of behaviors that we repeatedly execute without consciously trying to remember them (e.g., typing). Serial order learning is thus expressed in multiple memory systems [see Hurlstone, Hitch, & Baddeley, 2014; Page & Norris, 2009], and in the

context of autism it is now well established that explicit serial order memory is a source of difficulty, while implicit serial order learning is preserved [see Desaunay et al., 2020; Foti, De Crescenzo, Vivanti, Menghini, & Vicari, 2015, for reviews]. Although early studies by Hermelin and O'Connor had indicated relatively preserved memory for the serial order of short lists of words or pictures [Hermelin & O'Connor, 1970], over a dozen studies since then have shown that autistic children and adults demonstrate difficulties on digit-span and visuospatial span tasks, with group differences characterized by a medium effect size [Desaunay et al., 2020]. By contrast, implicit serial order learning, which is

typically assessed using Serial Reaction Time Tasks [SRT; Nissen & Bullemer, 1987], is generally preserved. SRT tasks require participants to respond as quickly as possible to stimuli that appear in repeating (or random) sequences of screen locations, and a meta-analysis by Foti et al. [2015] showed that out of seven such studies, only one [Mostofsky, Goldberg, Landa, & Denckla, 2000] indicated reduced learning in autistic as compared to typically developing participants. At least three studies since this review have further corroborated generally preserved SRT performance in autism [Zwart, Vissers, Kessles, & Maes, 2018; Zwart, Vissers, & Maes, 2018; Zwart, Vissers, van der Meij, Kessles, & Maes, 2017].

The dissociation between explicit and implicit serial order memory in autism is of considerable interest because a wealth of literature indicates that they play distinct roles in aspects of language development [e.g., Conti-Ramsden, Ullman, & Lum, 2015; Gathercole, Service, Hitch, Adams, & Martin, 1999; Mosse & Jarrold, 2008; Ullman, 2004]. In addition, there has been a long-standing debate concerning the etiologic overlap between specific language impairment (SLI) and autism [Boucher, 2012; Boucher & Anns, 2018; Ullman & Pullman, 2015; Williams, Botting, & Boucher, 2008]. A considerable literature now indicates that abnormalities in implicit serial order memory are a defining feature of SLI [Coady & Evans, 2008; Lum, Conti-Ramsden, Morgan, & Ullman, 2014; Obeid, Brooks, Powers, Gillespie-Lynch, & Lum, 2016]. By contrast, the literature mentioned above clearly suggests that at least implicit serial order learning processes are preserved in autism [see Boucher & Anns, 2018 in response to Ullman & Pullman, 2015]. Moreover, differences between SLI and autism are indicated by distinct patterns of performance on nonword repetition (NWR) tasks, which require participants to repeat back nonwords (e.g., blenkum) of varying lengths. Due to the requirement to retain the order of phonological units in memory, NWR performance relies, at least in part, on serial order learning processes [Page & Norris, 2009]. NWR is consistently impaired in SLI [Coady & Evans, 2008; Graf Estes, Evans, & Else-Quest, 2007] and to some extent also in autism accompanied by language impairments [e.g., Whitehouse, Barry, & Bishop, 2008; Williams, Payne, & Marshall, 2013]. However, the difficulty in autism is commensurate with verbal mental age, whereas in SLI the difficulties are more profound [see Nadig & Mulligan, 2017; Williams et al., 2013]. Moreover, the pattern of errors committed by children with SLI on NWR tasks is qualitatively different from those committed by autistic children with language impairments, possibly reflecting that serial order learning processes play different roles in language impairments in autism as compared to SLI [e.g., Whitehouse et al., 2008].

Unfortunately, studies examining the putative causes of language impairment in autism remain scarce, and even the few studies that do exist primarily include children who, despite significant language impairments, are functionally verbal and often have nonverbal abilities within the typical range. Approximately 30% of autistic children, however, are estimated to remain functionally nonverbal until at least secondary school [Tager-Flusberg & Kasari, 2013] and a very significant number also have severe learning disabilities and other comorbidities that impact upon adaptive functioning [Baird et al., 2006; Charman et al., 2011; O'Brien & Pearson, 2004]. This group is generally underrepresented in the literature [Russell et al., 2019], partly because of considerable methodological challenges in probing cognitive processes of interest. The overarching aim of the work reported here was therefore to develop an eye-tracking paradigm that would be suitable for probing serial order learning across the entire autism spectrum. Specifically, we took advantage of the fact that serial order memory is supported by domain-general learning processes that are expressed in different memory systems and across different modalities [see Hurlstone et al., 2014; Page & Norris, 2009]. The evidence in support of such domain-general processes stems from studies that have demonstrated equivalences in performance between visuospatial and verbal Hebb repetition paradigms that require participants to remember repeated sequences of dot-locations or verbal stimuli, respectively [Couture & Tremblay, 2006; A. J. Johnson, Dygacz, & Miles, 2017; Mosse & Jarrold, 2008]. Of particular interest in this literature is the observation that visuospatial Hebb repetition learning can be demonstrated through eye-tracking measures [Guerard, Saint-Aubin, Boucher, & Tremblay, 2011; Tremblay, Saint-Aubin, & Jalbert, 2006]. Specifically, participants demonstrate increasingly quicker gaze reaction times (including anticipatory fixations) to dot locations that form part of a repeating sequence. Moreover, Mosse and Jarrold [2008] have demonstrated that standard manual visuospatial Hebb repetition paradigms can be adapted to be very child-friendly. In their experiment, 5–6 year-old typically developing children were shown a frog appearing on five different lily-pads before being asked to try and reproduce the sequence. Unbeknownst to the children some of the sequences were repeated across trials and their reproduction of these sequences demonstrated the expected Hebb repetition effect. This effect was also correlated with the children's performance on a nonword repetition task, demonstrating the putative role of domain-general serial order learning processes in aspects of language development. Drawing on these studies by Tremblay et al. [2006] and Mosse and Jarrold [2008], we developed an eye-tracking task in which participants are simply asked to "*watch out for the bunny*" that

appears in repeating sequences of screen locations. In Experiment 1, we first examine whether adults with and without a diagnosis of ASD and no co-occurring language or intellectual difficulties demonstrate repetition learning on this task to establish the general feasibility of using such a paradigm to demonstrate the phenomenon. This was important, because, to the best of our knowledge, no previous study has examined repetition learning in the context of a passive viewing task that requires no overt behavioral responses from participants. Experiment 2 then applied the paradigm to a relatively large and heterogeneous sample of autistic and typically developing children to establish whether the paradigm can shed light on serial order learning processes across the autism spectrum.

Experiment 1: Adults
Methods

Participants. Twenty adults with a clinical diagnosis of Autism Spectrum Disorder (ASD) and 19 typically developing adults (TD) were recruited from an existing database at City, University of London, and through advertisement in the local area. Groups were matched in terms of age, gender, intellectual functioning [WAIS-III[UK]; Wechsler, 1997] and explicit phonological serial order memory, which was indexed by the Digit Span (DSp) subtest of the WAIS as well as the Nonword Repetition (NWR) subtest of the Comprehensive Test of Phonological Processing [CTOPP; Wagner, Torgesen, & Rashotte, 1999]. Although the CTOPP is standardized for use only up to the age of 24 years, pilot testing indicated that the NWR task is sufficiently difficult for adults to yield an informative range of raw-scores. Three ASD participants were subsequently excluded from all analyses due to poor quality data during the eye-tracking task (further details below). Descriptive statistics for the remaining 17 ASD and 19 TD participants are summarized in Table 1.

Participants in the ASD group had received their diagnosis through the UK's National Health Service in accordance with the DSM-IV-TR [American Psychiatric Association, 2000] criteria in force at the time of their diagnosis. ASD participants also met either the cutoff criteria on the revised algorithm for Module 4 of the Autism Diagnostic Observation Schedule (ADOS; see Hus & Lord, 2014], the cutoff score of 26 on the Autism Spectrum Questionnaire [AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Ruzich et al., 2015], or both. The ADOS was administered by a research reliable member of the research team to 13 of the 17 participants.[1] Eight participants met the cutoff criterion score of 6 or higher on the revised Social-Affect domain recommended by Hus and Lord [2014], with three participants receiving a score of 5 and two a score

of 4. Seven participants met the combined total algorithm criterion of 8, with four participants receiving a score of 7, and two participants receiving scores of 6 and 4, respectively. On the AQ, all except two participants had a score of 26 or higher, with the remaining two participants scoring 25 and 23, respectively. In the TD group all except three participants scored less than 26 on the AQ and all scored less than 32, which is the higher cutoff score recommended by Woodbury-Smith, Robinson, Wheelwright, and Baron-Cohen [2005] when screening for ASD in community-based rather than clinically referred samples. All TD participants confirmed that they did not have a family or personal history of psychiatric illness and that they did not take any medication or illicit substances.

Materials and design. The spatial serial order memory task ("*the Bunny Task*") consisted of a cartoon landscape showing a ring of eight bunny holes surrounded by some trees and shrubs. The task was presented in E-prime 2.0 on the 23″ screen of a Tobii TX300, which was set to record gaze coordinates throughout the task at a frequency of 120 Hz. The bunny holes measured approximately $1.9° \times 0.9°$ of visual angle at a viewing distance of 60 cm, with the ring of bunny holes extending 12.8° horizontally and 10.9° vertically. Each bunny hole could be animated for around 600 ms to show a white bunny jumping out of, and then disappearing back into the hole, pausing briefly midway with its arms over its head (see Fig. 1). On each of 20 experimental trials the bunny appeared in five locations at a rate of 1 every second, with an additional second separating successive trials. The background landscape including all bunny holes remained on the screen at all times. During the first five trials, the bunny appeared in a random sequence of locations. During the following 10 trials, the sequence shown in Figure 1 was repeated, and during the last five trials the locations were once again chosen randomly. The repeating sequence of locations was generated pseudo-randomly to avoid the bunny appearing in successive adjacent locations more than once, while ensuring that the path it took (i.e., the imaginary line connecting the five locations) crossed over once. The latter criterion was informed by studies suggesting that zero path crossings might lead to an unusual absence of primacy and recency effects that are otherwise a typical characteristic of serial order memory, while more frequent path crossings might hinder learning [Parmentier, Elford, & Maybery, 2005].

---

[1] Due to time constraints the ADOS could not be administered to the remaining four participants, who all provided records to confirm that their diagnosis had been established by suitably qualified clinicians.

Table 1. Descriptive Statistics for the ASD and TD Adults in Experiment 1

Measure	TD (16m, 3f)			ASD (14m, 3f)			Cohen's d
	M	SD	Range	M	SD	Range	
Age (years)	34.4	10.6	21.3–58.3	35.8	8.2	19.7–51.9	0.15
VIQ[a]	110.2	15.5	84–138	110.7	17.2	74–138	0.03
PIQ[b]	107.3	13.0	87–127	108.6	20.1	74–134	0.08
FIQ[c]	109.6	14.0	88–135	110.7	18.6	73–132	0.07
Digit Span	11.0	3.2	6–16	10.8	3.6	6–19	0.06
NW-Rep[d]	13.5	2.7	8–18	13.7	2.6	7–18	0.08
ADOS SA.[e]				7.2	3.1	4–14	–
ADOS RRB.[e]				1.9	1.1	0–4	–
ADOS Total[e]				9.2	3.4	4–16	–
AQ[f]	16.4	7.2	5–28	32.9	6.3	23–45	2.44

[a]Verbal IQ (WAIS-IIIUK).
[b]Performance IQ (WAIS-IIIUK).
[c]Full-scale IQ (WAIS-IIIUK).
[d]Nonword Repetition subtest of the CTOPP.
[e]Social Affect (SA), Restricted and Repetitive Behavior (RRB) and Total algorithm scores of the ADOS-2 [see Hus & Lord, 2014].
[f]Autism-Spectrum Quotient.
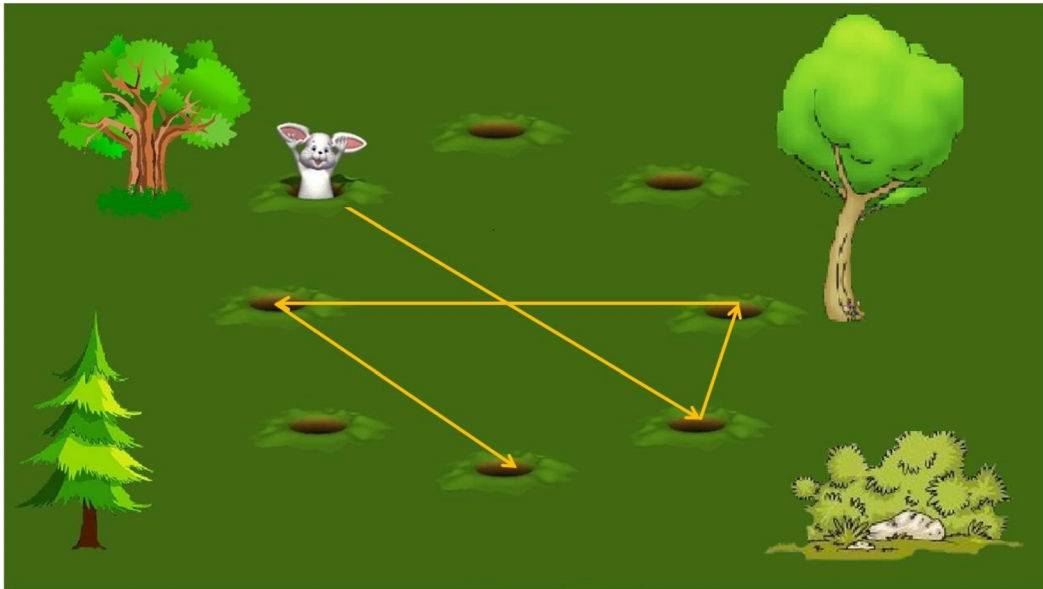
Figure 1. Cartoon environment used for the Bunny Task. The arrows show the sequence of locations the bunny appeared in during the repeating phase of Experiment 1.

Procedure. Participants were tested individually in a dedicated quiet room at City, University of London. They completed the Bunny Task and NWR as part of a 2.5-h visit during which they also completed either unrelated experiments and/or assessments such as the ADOS or WAIS if these were not already on file. Participants were informed that they would be asked to try a short eye-tracking task in which they simply had to "*watch out for a bunny*." We explained that the task was designed so that both adults and very young children could complete it, and that it would, therefore, be short and involve no further instructions other than to "*watch out for the bunny*."

Once participants gave their consent, an initial welcome screen showed the background scene and the bunny that participants were to watch out for. The five-point calibration procedure was then explained and carried out before the welcome slide was shown again until the participant confirmed that they were ready to start. The trial sequence, which lasted around 5 min, was then initiated.

Immediately after the task, participants were asked whether they had noticed any patterns in the locations that the bunny appeared in. Irrespective of whether they answered yes or no, they were told that, at a certain point in the task, the bunny repeated a certain sequence of five

locations. Participants were then asked to try to recall (or guess) this sequence by numbering relevant locations on printouts of the bunny holes from 1 to 5. At the end of the task, participants were debriefed and asked to complete the NWR task (the DSp was already on file for participants).

Data preparation and analysis. The raw gaze data were processed off-line using MATLAB® routines, which extracted details relating to all fixations that lasted a minimum of 100 ms.[2] Specifically, each of the eight bunny holes on the stimulus display was defined as a region of interest (ROI) and a time-stamp representing the onset of each bunny animation was added to the raw gaze data file during the experiment. Together with information about the onset time of each fixation, this allowed for the calculation of the latency between the onset of each bunny and the participant fixating the location where it appeared (hereafter the *target location*). For the analysis, these latencies were averaged across the five bunny fixations of each trial, and by fitting a straight line to these average latencies across the 10 repeating trial sequences, learning slopes could be derived for each participant. Importantly, because of the interval between one bunny in the sequence and the next, the fixation latencies could have negative values, indicating that fixations were anticipatory. The frequency of such anticipatory target fixations was therefore also derived for each trial, as were anticipatory fixations to other *in-sequence locations* and *out-sequence locations*. In-sequence locations included those locations that were part of the five-location sequence but were not the location at which the bunny was about to appear. Out-sequence locations were the three locations at which the bunny never appeared during the repeating set of trials.

Before the main analyses, the quality of the raw data were inspected for each participant by calculating (1) the proportion of missing raw data-points due to signal loss (i.e., the proportion of samples on which the eye-tracker could not establish gaze coordinates) and (2) the number of bunnies participants fixated throughout the task. Two ASD participants were excluded from further analysis at this stage because their raw data were characterized by more than 25% signal loss and one additional ASD participant was excluded for looking at fewer than 25% of the bunnies throughout the task. Preliminary analyses indicated that for these participants it was not possible to derive the dependent variables described above, and as such a data-driven approach was used to define exclusion criteria. The remaining participants did not differ in terms of signal loss (ASD: M = 5.6%, SD = 4.4%; TD: M = 5.4%, SD = 4.9%; t = 0.12, df = 34, p = 0.91, Cohen's

---

[2]Fixations were defined as successive gaze coordinates that fell within approximately 0.3° of visual angle over a period of at least 100 ms.

d = 0.04) or the percentage of bunnies that were fixated (ASD: M = 83.1%, SD = 14.7%; TD: M = 88.5%, SD = 15.8%; t = 1.05, df = 34, p = 0.32, Cohen's d = 0.35).

For all participants retained in the analyses, group differences in repetition learning were assessed using repeated measures ANOVAs and t-tests. In the ANOVAs, the Greenhouse–Geisser correction (GGC) was applied where the sphericity assumption was violated. Cohen's d and partial eta-squared (η_p^2) are reported as effect size measures and participant-level data are illustrated in figures to highlight important individual differences. Due to significant individual differences in the eye-tracking indices of repetition learning, it was not feasible to examine correlations between these indices and participant characteristics in this first experiment.

Results

Fixation durations and gaze latencies. Average fixation duration did not differ significantly between groups (ASD: M = 444 ms, SD = 88 ms; TD: M = 512 ms SD = 151 ms; t = 1.62, df = 34, p = 0.12, Cohen's d = 0.55), but the effect size was moderate with numerically longer average fixations in the TD group. Moreover, the data were characterized by marked individual differences in the TD group, which we will return to shortly.

Figure 2(A) illustrates the average gaze latencies to the onset of the bunnies for each of the 20 trials as a function of group. A 2 (group) × 10 (trials) repeated measures ANOVA across the 10 repeating trials demonstrated a main effect of group (F[1,34] = 4.87, p = 0.03, η_p^2 = 0.13) and a marginal linear trial × group interaction (F[1,34] = 3.50, p = 0.07, η_p^2 = 0.09). This interaction was the result of a significant linear decrease in gaze latencies in the TD (F[1,18] = 7.63, p = 0.013, η_p^2 = 0.29) but not the ASD group (F[1,16] = 0.33, p = 0.58, η_p^2 = 0.02).

Anticipatory fixation frequencies. Paralleling the latency data, an analysis of the frequency of anticipatory fixations to target locations (i.e., locations where the bunny was about to appear) also showed that only TD (F[1,18] = 7.95, p = 0.011, η_p^2 = 0.30) but not ASD participants (F[1,16] = 0.84, p = 0.37, η_p^2 = 0.05) anticipated the bunny increasingly more often across the repeating trials (Fig. 3(A)). Importantly, however, the data were characterized by considerable individual differences and closer inspection suggested that the group differences were primarily driven by a relatively small number of TD participants who exhibited a large number of anticipatory target fixations (Fig. 3(B)). These participants also demonstrated the steepest slopes in gaze latency decreases over the repeating trials (Fig. 2(B)) and the longest average fixation durations. This latter finding provides some context for understanding the greater variability and numerically longer average fixation durations in TD vis-à-vis ASD
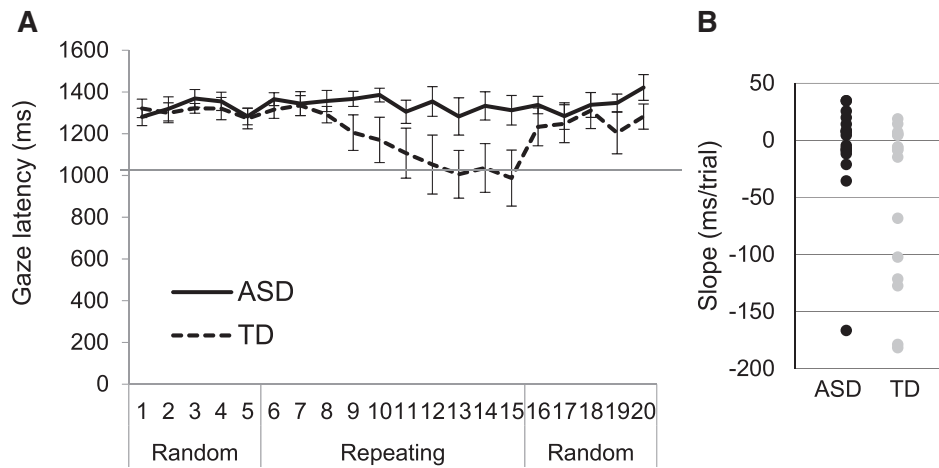
Figure 2. (A) Gaze latencies to bunny onset as a function of group. The solid line at 1000 ms indicates the onset of the bunny and error bars reflect ±1 standard error (SE). (B) Linear slopes for each participant across the 10 repeating trials.
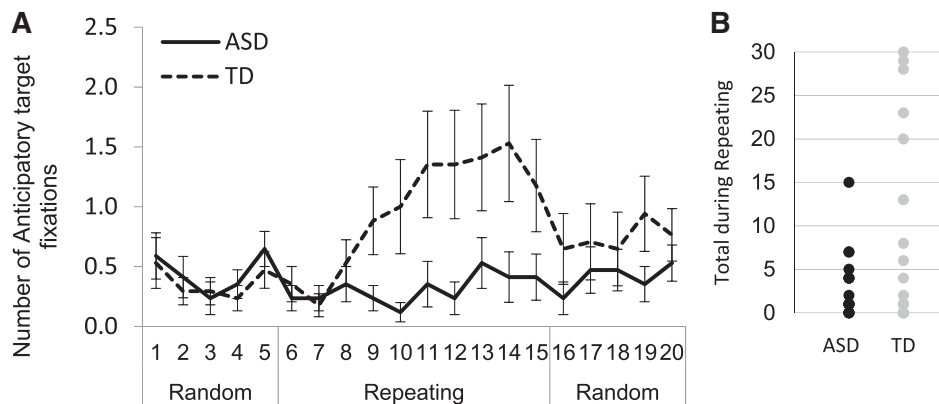
Figure 3. (A) Average number of anticipatory target fixations as a function of group. Error bars reflect ±1 standard error (SE). (B) Total number of anticipatory target fixations during the repeating trials for each participant.

groups noted above. Specifically, participants who orient their gaze to target locations increasingly quickly would look at them for longer or even wait for them to appear, thus increasing the duration of their fixations. In fact, learning slopes and average fixation durations were significantly correlated across both groups ($r = -0.56$, $p < 0.001$).

We next examined the frequencies of anticipatory fixations to *in-sequence* and *out-sequence* locations across the 20 trials. Increases in anticipatory fixations to *in-sequence* locations and/or decreases of anticipatory fixations to *out-sequence* locations would indicate that participants had learned in which five of the possible eight locations the bunny was likely to appear, even if they did not anticipate its exact location at exactly the right time. The data are shown in Figure 4. A 2 (group) × 20 (trial) repeated measures ANOVA of *in-sequence* fixations demonstrated a significant trial effect ($F[9.1,310.8] = 2.46$, $p = 0.001$, $\eta_p^2 = 0.07$, GCC). Although there was no main

effect of group ($F[1,34] = 0.04$, $p = 0.85$, $\eta_p^2 = 0.001$) or group × trial interaction ($F[9.1,310.8] = 1.08$, $p = 0.38$, $\eta_p^2 = 0.03$, GCC), as indicated by Figure 4, the increase in anticipatory *in-sequence* fixations was primarily driven by the TD group ($F[5.9,107.2] = 2.36$, $p = 0.036$, $\eta_p^2 = 0.12$, GCC); this trial effect was not significant in the ASD group ($F[7.8,125.2] = 1.16$, $p = 0.29$, GCC). By contrast, an equivalent analysis of anticipatory fixations to *out-sequence* fixations demonstrated a significant effect of group ($F[1,34] = 5.28$, $p = 0.03$, $\eta_p^2 = 0.13$), with an overall greater number of *out-sequence* fixations in the ASD ($M = 44.9$, SD = 30.9) than the TD group ($M = 25.5$, SD = 18.9).

Explicit awareness of sequences. Finally, we examined the extent to which participants in each group demonstrated explicit awareness and memory for the repeating sequence of locations. Only 3 (18%) ASD and 7 (37%) TD participants said that they had noticed a
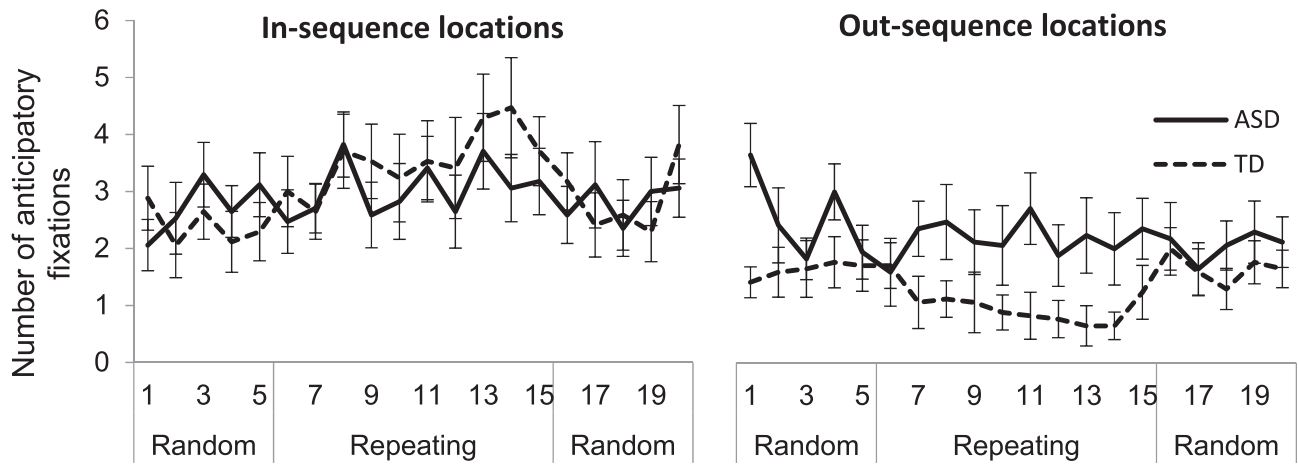
Figure 4. Average frequencies of anticipatory fixations to In-sequence and Out-sequence locations as a function of group. Error bars reflect ±1 standard error (SE).

pattern in the sequence of bunny locations, which represents a nonsignificant difference between the groups ($x^2[1] = 1.65$, $p = 0.20$). When attempting to reproduce the sequence of locations on paper, the TD group ($M = 3.8$, SD = 0.78) indicated a significantly greater number of the correct locations than the ASD group ($M = 3.0$, SD = 0.87); $t = 2.87$, df = 34, $p = 0.007$; Cohen's $d = 0.97$). However, within the locations indicated there were no group differences in terms of identifying locations in their correct serial order (TD: $M = 1.4$, SD = 1.6; ASD: $M = 1.2$, SD = 0.8; $t = 0.44$, df = 34, $p = 0.67$, Cohen's $d = 0.16$). Importantly, across both groups, participants who indicated awareness of a pattern in the bunny locations demonstrated more anticipatory fixations ($t = 2.00$, df = 11.9, $p = 0.07$, Cohen's $d = 0.81$) and steeper learning slopes ($t = 2.95$, df = 11.4, $p = 0.013$, Cohen's $d = 1.22$) than those who indicated no awareness.

Interim Discussion

The data from Experiment 1 demonstrate that the Bunny Task elicits a robust repetition learning effect in at least some TD participants as indicated by gaze latency decreases to bunny locations on the one hand and increases in anticipatory fixations to bunny locations across repeating trial sequences on the other. Somewhat surprisingly the ASD group demonstrated no such repetition learning effect, which may have been related to difficulties in inhibiting fixating the three out-sequence locations during the repeating trial sequences.

An important characteristic of the data in Experiment 1 was that the repetition learning effect observed in the TD group was characterized by considerable individual differences and a discontinuous distribution of indices of learning (i.e., the slope of gaze latency decreases across trials and the overall number of anticipatory target

fixations). Closer inspection of the data (see Fig. 2) revealed that group differences were primarily driven by a relatively small number of TD adults who demonstrated the steepest gaze latency decreases and the greatest number of anticipatory fixations over the course of the repeating trial sequences. These participants were also those who indicated that they had become aware of a pattern in the sequence of bunny locations. In Experiment 2, we therefore sought to simplify the Bunny Task for use with children by removing the initial phase of random location sequences.

Experiment 2: Children
Methods

Participants. Eighty-six children with a diagnosis of Autism Spectrum Disorder (ASD) and 44 typically developing children (TD) participated in the study. Seventy ASD and 28 TD children were recruited from mainstream and special educational needs schools in the south-west of England (UK sample), while the remaining 16 ASD and 16 TD children were enrolled in the Autism Phenome Project (APP) at the University of California, Davis, MIND Institute (US sample). The APP is a longitudinal study which began in 2006 when participants were 24 to 44 months old. Participants completed the Bunny Task (described below) as part of their Time Point 4 visits, which are conducted when participants are in middle childhood (age 9–12 years). ASD children recruited in the United Kingdom had all received their diagnosis through local health services and had a statement of Special Educational Needs that mentioned autism as the child's primary need for adjustment. Parents were also asked to complete the Social Communication Questionnaire [SCQ; Rutter, Bailey, & Lord, 2003] to further corroborate the diagnosis. The response rate was unfortunately low,

but 32 of 37 SCQs that were returned demonstrated scores above the recommended cutoff score of 15 (M = 21.1, SD = 6.27), with the five remaining participants receiving scores of 12 or higher. In the US sample, all participants with ASD had their diagnosis confirmed by a licensed clinician at the UC Davis MIND Institute during initial enrolment in the APP, when all met criteria for ASD based on the NIH Collaborative Programs of Excellence in Autism standards. This means that they met the criteria for ASD on the Autism Diagnostic Observation Schedule [ADOS-2; Lord et al., 2000], and met criteria for autism on either the Social or Communication subscale of the Autism Diagnostic Interview—Revised [Lord, Rutter, & Le Couteur, 1994] while also being within two points of criterion on the other subscale. The ADOS-2 was readministered at Time Point 4 of the APP when the bunny task was also administered and these ADOS data are included in Table 2 (one child no longer met criteria but their data were retained in all analyses to preserve the representativeness of our sample). Teachers and/or parents of TD children in both the UK and US samples confirmed that no concerns had been raised about their development or educational progress. For 14 of the children in the US sample, parents also returned the SCQ, which confirmed below cutoff scores in all cases. In the UK sample, only four parents of TD children returned the SCQ and all confirmed below cutoff scores but due to the poor response rate these data are not included in Table 2.

To characterize the children's nonverbal and verbal abilities, children in the UK sample were asked to complete the appropriate version of the Raven's Progressive Matrices [RPM; Raven, 1956], and the Vocabulary and Similarities subtests of the Wechsler Intelligence Scale for Children [WISC-III[UK]; Wechsler, 1991]. Their phonological serial order memory was assessed through the Digit-Span (DSp) and Nonword Repetition (NWR) subtest of the Comprehensive Test of Phonological Processing [CTOPP; Wagner et al., 1999], which were averaged to derive a single index of explicit phonological serial order memory (pSOM; DSp and NWR were highly correlated, r = 0.90, $p < 0.001$). Due to unavailability for all assessment sessions, Raven's data were missing for four ASD and two TD participants, WISC data for three ASD and four TD children and the DSp and NWR for two ASD and four TD participants. Children in the US sample completed the Differential Abilities Scale-II [DAS-II. Elliot, 2007], which provides a measure of general conceptual ability, as well as verbal and nonverbal subscales. A measure of pSOM was not available for this group. As shown in Table 2, both in the UK and US samples, ASD and TD groups did not differ in age (UK: t = 0.85, df = 94, p = 0.40, Cohen's d = 0.20; US: t = 1.45, df = 30, p = 0.16, Cohen's d = 0.52) but the ASD groups demonstrated significantly lower verbal abilities (UK: t = 6.83, df = 89,

$p < 0.001$ Cohen's d = 1.65; US: t = 3.39, df = 16.4, p = 0.002, Cohen's d = 1.20) and nonverbal abilities (UK: t = 4.28, df = 90, $p < 0.001$, Cohen's d = 1.09; US: t = 3.04, df = 30, p = 0.005, Cohen's d = 1.07) than the TD groups. In the UK sample the ASD group also demonstrated significantly lower pSOM scores than the TD comparison group (t = 5.94, df = 77.15, $p < 0.001$, Cohen's d = 1.20). It is worth noting that, in keeping with the literature, this difference was no longer significant when controlling for participant's VIQ ($F[2,87]$ = 0.78, p = 0.38, η_p^2 = 0.01), which was correlated with pSOM (r = 0.65, $p < 0.001$).

Importantly, 37 ASD children in the UK sample and five in the US sample had VIQs below 70 and of those 27 (25 UK; 2 US) could be described as minimally verbal because their performance on the WISC or DAS-II was too low to derive a verbal IQ. For 15 children in the UK sample (including 13 of the minimally verbal children) the Raven's also proved too difficult to derive a nonverbal IQ[3] and 10 children (all minimally verbal) could not meet the demands of the DSp and NWR tasks (they received a pSOM score of 0). Following preliminary analyses of the eye-tracking data (described shortly), 31 ASD children (29 in the UK sample; two in the US sample) and four TD children (three in the UK sample; one in the US sample) needed to be excluded from further analyses because relevant fixation frequencies and latencies could not be derived reliably. Table 2, therefore, summarizes descriptive statistics separately for all participants who were initially recruited and those who were ultimately retained in all analyses. Children who were excluded were significantly younger (t = 2.87, df = 128, p = 0.005, Cohen's d = 0.52) and had significantly lower verbal (t = 5.26, df = 121, $p < 0.001$, Cohen's d = 1.08) and nonverbal IQs (t = 3.69, df = 122, $p < 0.001$, Cohen's d = 0.76) than those who were included. Notably, 23 of the excluded children had VIQs below 70 and 18 of those children were minimally verbal. Figure 5 illustrates this relationship between VIQ and participant exclusion in detail and shows that 46% of children with IQs below 70 were ultimately retained in all analyses, compared to 76% of participants with IQs above 70. Of the 27 minimally verbal children, however, 18 (67%) ultimately needed to be excluded, who were significantly younger (t = 2.54, df = 25, p = 0.018, Cohen's d = 1.12) than the nine minimally verbal children who were included in subsequent analyses (see Table 3 for details). We will return to the implications of these findings in the discussion.

Materials and design. The materials and design for the Bunny Task were identical to those described for Experiment 1 with the following exceptions. First, we did not attempt to assess participants' explicit awareness of the

---

[3]For the calculation of group averages of IQs we assumed IQs of 45 for children for whom no actual IQ could be derived.

Table 2. Descriptive Statistics for the ASD and TD Groups of Experiment 2

| | UK sample | | | | | | US sample | | | | | |
| | ASD | | | TD | | | ASD | | | TD | | |
Measure	n	M (SD)	Range	n	M (SD)	Range	n	M (SD)	Range	n	M (SD)	Range
All participants												
Male:female	56:14			17:11			15:1			13:3		
Age (years)	70	12.3 (3.5)	4.9–18.6	23	11.6 (3.1)	3.8–14.7	16	11.6 (0.9)	9.9–12.9	16	11.2 (0.5)	10.2–12.4
VIQ[a]	67	68.5 (23.0)	45–130	24	105.3 (21.7)	71–155	16	89.2 (32.4)	45–150	16	117.3 (7.0)	102–128
NVIQ[b]	66	76.5 (21.3)	45–120	26	95.7 (13.1)	75–130	16	90.5 (17.7)	60–115	16	108.7 (16.3)	90–143
pSOM[c]	68	9.5 (4.8)	0–17	24	8.5 (2.5)	8.5–18.5	–			–		–
SCQ[d]	36	21.1 (6.3)	12–34	–	–	–	14	19.2 (7.6)	3–29	14	1.9 (2.5)	0–10
ADOS CSC[e]	–			–	–	–	16	7.2 (2.0)	3–10	–	–	–
Retained for ET analysis												
Male:female	33:8			16:9			13:1			12:3		
Age (years)	41	13.4 (2.9)	7.5–18.6	25	11.7 (3.2)	3.8–14.7	14	11.7 (0.9)	9.9–12.9	15	11.2 (0.5)	10.2–12.4
VIQ[a]	41	76.8 (20.7)	45–130	21	105.5 (23.2)	71–155	14	90.9 (32.1)	45–150	15	117.4 (7.2)	102–128
NVIQ[b]	40	80.7 (18.9)	45–120	24	95.7 (13.6)	75–130	14	91.9 (16.7)	63–115	15	109.3 (16.7)	90–143
pSOM[c]	41	11.2 (3.2)	0–17	21	14.4 (2.3)	10.5–18.5	–			–		–
SCQ[d]	22	22.1 (7.1)	12–34	–	–	–	13	18.6 (7.6)	3–29	13	1.9 (2.6)	0–10
ADOS CSC[e]	–			–	–	–	14	7.2 (2.0)	3–10	–	–	–

Note. The top half of the table summarizes the data for all participants who were initially recruited; the bottom half of those who were retained in all analyses.
[a]Verbal IQ as assessed using the Vocabulary and Similarities subtests of the WAIS-III[UK] in the UK sample and the DAS verbal subscale in the U.S. sample.
[b]Nonverbal IQ as assessed using the Raven's colored and standard progressive Matrices in the UK sample and the DAS nonverbal subscale in the U.S. sample.
[c]pSOM was assessed as the average raw score of the Digit Span and Nonword Repetition subtests of the CTOPP.
[d]Social Communication Questionnaire.
[e]Autism Diagnostic Observation Schedule Calibrated Severity Scores.
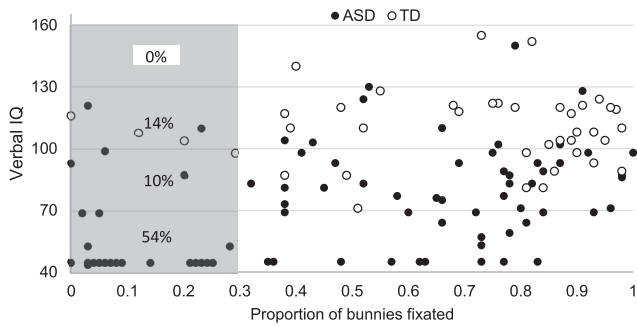
Figure 5. Scatter plot illustrating the relationship between verbal IQ and the proportion of bunnies participants fixated. The shaded area indicates the criterion range for excluding participants from further analysis and the % values in the shaded area indicate the percentage of participants within each 30 point IQ band who were excluded.

bunny sequence as it is not clear that such a measure would yield reliable results in this heterogeneous sample [see Mosse & Jarrold, 2008; Mosse & Jarrold, 2010]. Second, random bunny sequences were removed and participants simply watched the same five-location sequence on each of 19 trials.[4] The sequence was randomly generated as explained for Experiment 1. For pilot-purposes, in the UK sample only, five additional trials were included in which two randomly selected locations in the sequence (though never the first location) swapped position. These trials were interspersed with the repeating sequences during the second half of the experiment (trials 13, 16, 18, 22, and 24) and served to determine whether gaze latencies to unexpected bunny appearances might serve a useful additional index of learning. These trials were not included for the US sample (where testing started after some preliminary data were available from the UK sample) because we were concerned that the inclusion of such trials might interfere with learning during later stages of the experiment. The only other difference to the experimental setup of Experiment 1 was the inclusion of some additional introductory slides at the beginning of the task to make the purpose of the "game" clear for children (see procedure section below for details). Finally, because testing of the UK sample took place in the children's schools, a portable Tobii X1 Light Eye-Tracker (Tobii Technology), attached to the screen of a 14″ laptop monitor, was used to monitor gaze fixations throughout the task. In the United States, testing was performed at the UC Davis MIND Institute, where a Tobii 1750 LCD binocular tracker (Tobii Technology) was used for the experiment.

---

[4]Due to a programming error, four ASD and two TD children were presented with a twentieth repeating trial that was simply ignored in the analysis as it had no bearing on the other 19 trials.

Procedure. After obtaining written consent from parents, each child in the UK sample was tested individually over two or three 10–20 min sessions in a quiet room at the child's school. In the United States, children visited the UC Davis MIND Institute several times for their Time Point 4 (middle childhood) testing. The Bunny Task and DAS-II were administered, among a larger battery of behavioral tasks and assessments, during these visits. The RPM and WISC subtests in the UK, and the DAS in the US, were administered according to the standard instructions provided by the relevant manuals. For the Bunny Task, children were asked to help the experimenter "*watch out for a bunny,*" which would be jumping out of different holes on the screen. An introduction slide showed the bunny disappearing into a bunny hole next to a big sign that welcomed the child to the "game." The experimenter explained that this was the bunny the child should "*watch out for*" and that, in a moment, there would be several bunny holes on the screen that the bunny could pop out of. The experimenter next explained that the gadget below the screen (the Tobii) would keep track of how well the child watched out for the bunny and that this gadget needed to be switched on before the task could begin. A standard five-point calibration procedure was then administered and after reminding the child one more time to simply keep "*watching out for the bunny*" the experimental trials were launched. No mention was made of the fact that the bunny would be repeating a particular sequence. The total duration of the task, following calibration, was approximately 4 min and 30 s and the experimenter was present throughout and provided encouragement for children to keep "watching out for the bunny" if they disengaged from the screen.

Data preparation and analysis. Data concerning fixation latencies and frequencies were derived in the same way as described for Experiment 1 and data quality was also examined in the same way (see "data preparation and analysis section" of Experiment 1 for details). Preliminary analyses indicated that for participants who fixated fewer than 30% of the bunnies it was not possible to derive fixation latencies and frequencies and therefore 31 ASD and four TD participants were excluded at this stage. On average, this excluded group demonstrated signal loss on 47% (SD = 19%) of the raw data samples and fixations were recorded for only 11% of all bunnies. For participants who were retained in all analyses (55 ASD and 40 TD), there were no group differences in the amount of signal loss (ASD: $M = 15\%$, SD = 10%; TD: $M = 12\%$, SD = 14%; $t = 1.36$, df = 93, $p = 0.18$, Cohen's $d = 0.25$) and although somewhat fewer bunnies were fixated by the ASD ($M = 68\%$, SD = 19%) than the TD group ($M = 76\%$, SD = 19%; $t = 2.11$, df = 93, $p = 0.038$, Cohen's

Table 3. Descriptive Statistics for the Minimally Verbal Autistic Children Who Either Were to Be Excluded From the Eye-Tracking Analysis or Could Be Included

Measure	Excluded			Included			
	n	M (SD)	Range	n	Included M (SD)	Range	Cohen's d
Male:female	14:4			9:0			
Age (years)	18	9.7 (3.8)	5.0–17.7	9	13.3 (2.4)	9.3–16.9	1.12
NVIQ[a]	17	60.5 (23.4)	45–115	9	61.0 (9.9)	45–75	0.03
pSOM[b]	18	4.4 (5.1)	0–13	7	7.5 (3.9}	0–12	0.68
SCQ[c]	11	20.6 (4.8)	14–29	7	23.1 (5.9)	14–31	0.47
Bunnies fixated	18	12%	0%–25%	9	59%	35%–83%	3.48
Signal loss[d]	18	53%	27%–78%	9	24%	8%–50%	1.83

[a]Nonverbal IQ as assessed using the Raven's colored and standard progressive Matrices in the UK sample and the DAS nonverbal subscale in the U.S. sample.

[b]pSOM was assessed as the average raw score of the Digit Span and Nonword Repetition subtests of the CTOPP.

[c]Social Communication Questionnaire.

[d]Signal loss refers to the percentage of raw eye-tracking data samples during which the eye-tracker could not establish gaze coordinates (e.g., because of inattention to the screen or blinks).
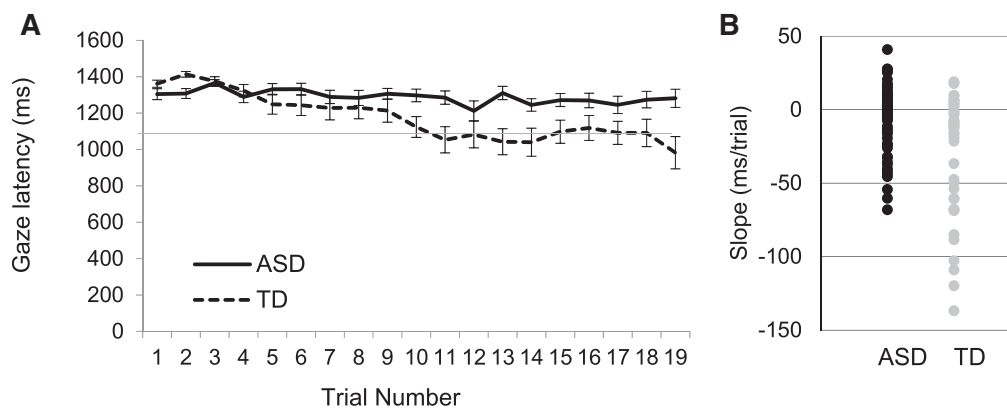
Figure 6. (A) Gaze latencies to bunny onsets as a function of group. The solid line at 1000 ms indicates the onset of the bunny and error bars reflect ±1 standard error (SE). (B) Linear slopes for each participant across the 19 trials.

$d = 0.42$) this did not significantly impact on the ability to derive relevant gaze latencies.

Following the initial data screening, a set of preliminary analyses were carried out to compare the results in the US and UK samples. The pattern of results in each subsample demonstrated the same pattern of effects described below, both when considering only the first 12 trials of the task that were identical across the two sites, and when considering all 19 repeating trial sequences. The data were therefore collapsed across the two sites for the analysis of the repeating trial sequences, while responses to the sequence violations are presented separately for the UK sample only.

Results

Fixation durations and gaze latencies. In keeping with the results of Experiment 1, groups differed in terms of average fixation durations with significantly longer fixations in TD ($M = 424$ ms, $SD = 94$ ms) compared to ASD participants ($M = 365$ ms, $SD = 56$ ms; $t = 3.49$, df = 59.1, $p = 0.001$, Cohen's $d = 0.76$). Figure 6(A) illustrates the average gaze latencies to the bunny onsets for each of the 19 repeating trial sequences as a function of group. In the TD group these latencies could not be derived for 21 out of the 760 trials in total across all participants (i.e., 3%) because of a lack of fixations to any of the five bunnies in those trials. In the ASD group the same was true for 38 out of the 1045 trials in total (i.e., 4%). To retain all participants in the analysis, these missing values were interpolated with the average latency of the two adjacent trials. A 2 (group) × 19 (trials) repeated measures ANOVA of the data demonstrated main effects of group ($F[1,93] = 5.87$, $p = 0.017$, $\eta_p^2 = 0.06$) and trial ($F[9.36,870.54] = 8.63$, $p < 0.001$, $\eta_p^2 = 0.09$, GGC) that were further qualified by a significant trial × group interaction ($F[9.36, 870.54] = 4.59$, $p < 0.001$,
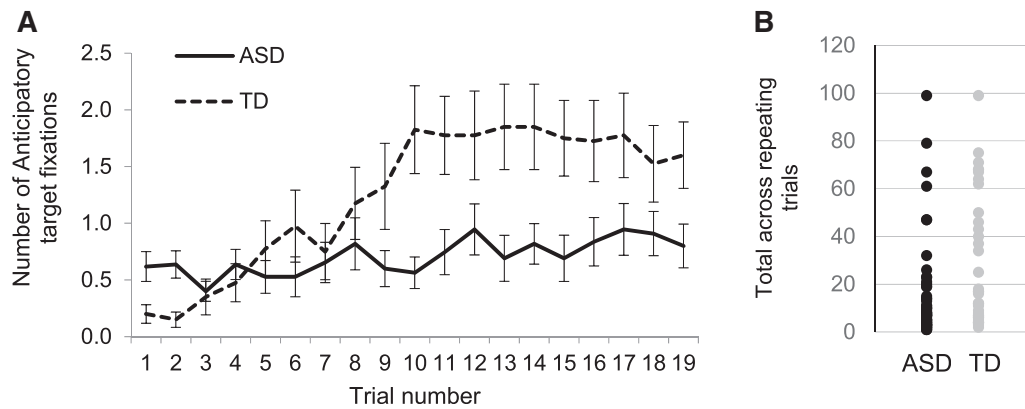
Figure 7. (A) Average number of anticipatory target fixations as a function of group. Error bars reflect ±1 standard error (SE). (B) Total number of anticipatory target fixations during the 19 repeating trials for each participant.
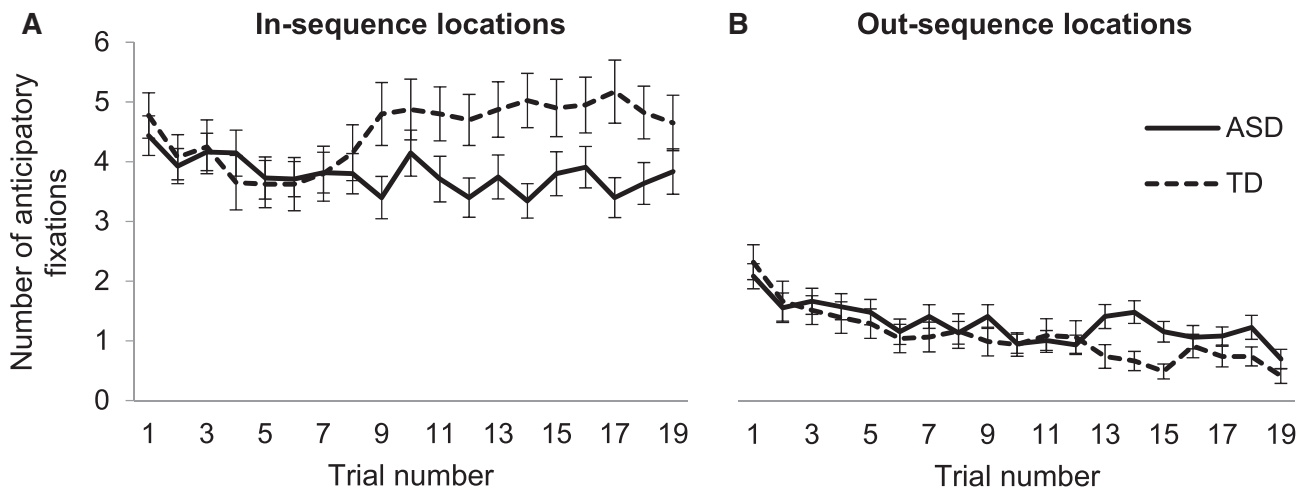
Figure 8. Average frequencies of anticipatory fixations to In-sequence and Out-sequence locations as a function of group. Error bars reflect ±1 standard error (SE).

$\eta_p^2 = 0.05$, GGC). Within-group follow up analyses showed that this interaction was the result of a significant trial effect in the TD group ($F[6.83,266.24] = 8.02$, $p < 0.001$, $\eta_p^2 = 0.17$, GGC) but not the ASD group $F(9.47,511.18) = 1.41$, $p = 0.12$, $\eta_p^2 = 0.03$, GGC).

Anticipatory fixation frequencies. The frequency of anticipatory fixations to target locations is illustrated in Figure 7(A) and demonstrated a significant increase in anticipatory target fixations across trials in the TD ($F[6.33,246.94] = 6.92$, $p < 0.001$, $\eta_p^2 = 0.15$, GGC) but not the ASD group ($F[7.42,400.82] = 1.47$, $p = 0.17$, $\eta_p^2 = 0.03$, GGC). It is worth noting that, as in Experiment 1, the data were characterized by considerable individual differences as indicated by the learning slopes calculated across the first 12 trials, and the total number of anticipatory target fixations (see Figs. 6(B) and 7(B)). Similar to Experiment 1, the learning slope was

significantly correlated with individual differences in average fixation durations across both groups ($r = -0.42$, $p < 0.001$).

Figure 8 illustrates the frequency of anticipatory in-sequence and out-sequence fixations. The data for in-sequence fixations were characterized by a significant trial by group interaction ($F[12.77, 1187.99] = 2.78$, $p = 0.001$, $\eta_p^2 = 0.03$, GGC), which was the result of significant increases in such fixations across trials in the TD ($F[9.30, 362.56] = 2.56$, $p = 0.007$, $\eta_p^2 = 0.06$, GGC) but not the ASD group ($F[12.92, 697.77] = 1.44$, $p = 0.14$, $\eta_p^2 = 0.03$, GGC). By contrast, anticipatory fixations to out-sequence locations decreased significantly across trials for both the ASD ($F[12.28, 663.24] = 4.08$, $p < 0.001$, $\eta_p^2 = 0.07$, GGC) and TD group ($F[9.20, 358.91] = 5.23$, $p = 0.001$, $\eta_p^2 = 0.12$, GGC), with no group by trial interaction ($F[12.61, 1173.42] = 1.39$, $p = 0.16$, $\eta_p^2 = 0.02$, GGC).
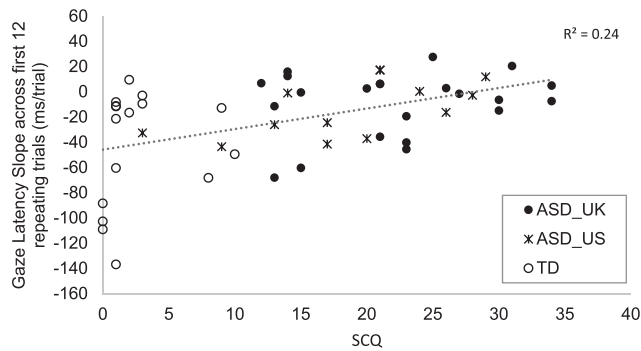
Figure 9. Scatter plot illustrating the association between SCQ scores and the slope of the gaze latency change over the first 12 repeating trial sequences. The regression line represents the association across all participants.

Effect of sequence violations. Next, we examined responses to the sequence violations in the UK sample. For this analysis, we compared the average gaze latencies to the expected bunnies in the last seven repeating trial sequences (those interspersed with the violated trial sequences) with the average gaze latencies to the unexpected bunnies that appeared in the swapped positions during the violated trials. One TD and five ASD participants needed to be excluded from this analysis because relevant "swapped" bunnies were not fixated. A 2 (group) × 2 (expected vs. unexpected) ANOVA of the available data demonstrated slower gaze latencies to the unexpected (M = 1324 ms; SD = 157 ms) compared to the expected (M = 1266 ms; SD = 219 ms) bunnies (F[1, 58] = 7.61, p = 0.008, η_p^2 = 0.12). The ASD group also demonstrated marginally slower response times than the TD group overall (F[1,58] = 3.75, p = 0.058, η_p^2 = 0.06). However, there was no 2-way interaction (F[1,58] = 0.21, p = 0.65, η_p^2 = 0.00) and within each group the effect of the sequence violation was of a very similar effect size (ASD Cohen's d = 0.32; TD Cohen's d = 0.31).

Correlations between eye-tracking indices of learning and participant characteristics. There were no significant correlations between eye-tracking indices of repetition learning (i.e., the slope in gaze latencies and the number of anticipatory target fixations) and children's age, verbal IQ and nonverbal IQ (rs < 0.18). In the UK data, there was also no correlation between the learning indices and children's performance on the NWR and Digit Span measures (rs < 0.24). However, there was a significant correlation between gaze latency slopes and SCQ scores (r = 0.49, n = 51, p < 0.001) and between anticipatory target fixation frequencies and SCQ scores (r = −0.36, n = 51, p = 0.01). These correlations held when considering only the ASD group alone (for slope: r = 0.36, n = 35, p = 0.035; for anticipatory fixations r = −0.35, n = 35, p = 0.038). For illustrative purposes, Fig. 9 shows

the relationship between the SCQ and the learning slope measure (the learning slope was highly correlated with the number of anticipatory target fixations r = −0.77, n = 95, p < 0.001).

General Discussion

The overarching aim of the two experiments presented above was to develop a new eye-tracking paradigm that could shed light on serial order memory processes across the entire autism spectrum, including particularly those individuals with significant language impairments and/or learning disabilities who remain grossly underrepresented in the literature. Across two experiments the Bunny Task elicited reliable repetition learning effects in typically developing adults and children, extending previous studies which have demonstrated reliable repetition learning effects using implicit eye-tracking measures [Tremblay et al., 2006]. Specifically, gaze latencies to bunnies appearing in repeating sequences of screen locations reliably decreased across trials and participants also demonstrated increased numbers of anticipatory fixations to the locations at which the bunny was about to appear. Somewhat surprisingly, across both experiments, there was a marked absence of these repetition learning effects in autistic adults (Experiment 1) and children (Experiment 2). Importantly, approximately half of the children who had significant language impairments and/or learning disabilities (i.e., IQs below 70) provided reliable data and could, therefore, be included in all analyses although 67% of the minimally verbal children who had the most significant difficulties needed to be excluded. In what follows, we first consider the implications of this last finding, before turning to possible explanations for the marked group differences in repetition learning.

As noted in the introduction, our knowledge of autism currently stems primarily from studies involving autistic individuals who have either no or only relatively mild language impairments and/or learning disabilities. Studies rarely include the estimated 30% of autistic individuals who are minimally verbal and have profound intellectual disabilities. A recent snap-shot review of all studies published in autism-specific journals in 2016, for example, indicated that of 100,245 autistic participants included across 301 studies, only 6% had intellectual disabilities and only 2% were described as minimally verbal [Russell et al., 2019]. There is, therefore, an urgent need for studies that include more representative samples of autistic individuals, particularly to shed light on processes that may play an important role in the heterogeneity in language development and intellectual functioning across the autism spectrum [e.g.,Boucher, Mayes, & Bigham, 2012; Bigham, Boucher, Mayes, & Anns, 2010; Tager-Flusberg et al., 2017]. Eye-tracking methods appear

to be ideally suited for this purpose because of their non-invasive nature and the flexibility with which they can be adapted to different experimental paradigms [Johnson, Lum, Rinehart, & Fielding, 2016; Lai et al., 2013]. The method is already extensively used in autism research to probe aspects of social cognition [Falck-Ytter, Bölte, & Gredebäck, 2013] and attentional control (Johnson et al., 2016) and some recent studies have begun to extend eye-tracking methods to study language processes in minimally verbal autistic children, adolescents and adults [Coderre et al., 2019; Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016]. Skwerer et al. [2016] for instance, examined receptive language in a group of 19 minimally verbal autistic children and adolescents using a preferential looking task in which participants were shown 84 picture pairs on a screen before hearing a spoken word describing one of the pictures. Eye-tracking indices were generally correlated with formal measures of participants' receptive vocabulary, but there were also substantial individual differences in data quality. Specifically, participants demonstrated fixations to one of the pictures on between 36% and 97% of the trials, and across all participants nearly half of the trials did not demonstrate any valid fixation data. The authors, therefore, acknowledged that eye-tracking methods have their limitations when used in studies involving minimally verbal autistic children, and the current findings corroborate this caution. In Experiment 2, participants fixated anywhere between none and all of the bunnies during the task and closer inspection of individual differences showed that data reliability was associated with IQ. Because of the nature of the Bunny task, around half of the children with IQs below 70 ultimately had to be excluded from the analyses because the relevant indices of repetition learning could not be derived from the available data. When considering only minimally verbal children 67% were ultimately excluded, who were on average younger than those who were retained. While this highlights certain limitations in the application of eye-tracking methods to research involving autistic individuals with complex needs, the current findings along with those of Skwerer et al. [2016] nevertheless also illustrate that reliable data can be acquired from a significant number of autistic individuals who have hitherto remained underrepresented in the literature.

An important question that is raised by current attempts to use eye-tracking technology to gain meaningful insights into cognitive strengths and weaknesses in autistic individuals with more complex needs, is how to maximize the utility of this technology to obtain reliable data from a larger majority of participants. In developing the Bunny Task, one of our main considerations was to keep the task as short as possible to minimize demands on extended periods of sustained attention. We still believe that this is important, particularly for very passive

tasks. However, future studies employing this type of paradigm should consider repeating the procedure over several testing sessions, perhaps using different animated sequences informed by parents to ensure the materials are engaging (for instance, special interest stimuli could be used, or characters from a favorite cartoon). Although such an approach would be more costly and time-consuming (note to funding bodies!), it should lead to greater inclusion of participants by offering multiple opportunities to engage with the task. An additional benefit would be that the test–retest reliability of the paradigm could be assessed. Another option for increasing task engagement for paradigms such as the Bunny Task is to make the task more interactive. In the process of developing the Bunny Task, we had also trialed two additional interactive versions that we thought might yield better engagement. One version was set up to be gaze-contingent such that bunnies would appear only when participants had fixated the preceding bunny. The other version presented dirt-piles that occasionally blocked one of the rabbit holes and participants had to remove the pile by touching it to help the bunny "continue playing in the garden." Neither of these manipulations was successful and during pretesting it quickly became apparent that disrupting the rhythm of the animated sequence hindered repetition learning. Thus, at least for repetition learning paradigms, either a fully passive viewing task such as here, or a fully interactive manual task such as standard serial reaction time tasks appear to be the most appropriate format.

Turning to the group differences in the repetition learning effects that were evident across both experiments, it seems unlikely that these were the result of the ASD participants not engaging with the task demands. For one, across both experiments, there were no group differences in gaze latencies to bunny onsets during the initial trials and throughout the task, both groups fixated a similar number of bunnies (setting aside participants who were excluded from the analyses). In Experiment 2, similar decreases in anticipatory fixations to out-sequence locations were evident in both groups, which suggests that ASD participants acquired some knowledge about where the bunny was likely to appear but less about the specific serial ordering of these bunny locations. This parallels findings from explicit serial order memory tasks in both the verbal [Poirier, Martin, Gaigg, & Bowler, 2011] and visuospatial domain [Bowler, Poirier, Martin, & Gaigg, 2016]. More generally this finding shows that participants were engaged with the task demands. In Experiment 2, this is further corroborated by the fact that both groups demonstrated similar gaze-latency slowing to the bunnies that appeared in unexpected locations.

Another possibility for the group differences that seems unlikely is that it reflects a general ocular motor difficulty

in ASD (see Johnson et al., 2016 for a comprehensive review of this literature). Although there are some indications in the literature of such difficulties, these tend to be evident primarily in smooth pursuit tasks or tasks that require the inhibition or precise control of saccadic movements. There is little evidence for difficulties in fixating simple visual targets as required in the current task, particularly when no concurrent distracters are present. Moreover, although group differences were apparent in the current experiments in the average fixation durations, these differences appeared to be a reflection of the group differences in repetition learning since average gaze durations were correlated with individual differences in learning slopes.

If general group differences in task engagement or ocular motor control do not account for the absence of repetition learning effects in autism in the current experiments, it is important to consider why these observations might conflict with the consensus that such learning is generally preserved. One possibility is that the Bunny Task taps explicit rather than implicit serial order learning processes. Somewhat paradoxically, previous studies examining implicit serial order learning using SRT tasks in autism, have gone to great lengths to minimize the influence of explicit awareness to control for the possibility that autistic participants might compensate for implicit learning difficulties with explicit strategies [see Foti et al., 2015 for a systematic review; see also Zwart et al., 2017]. However, as outlined in the introduction, if anything, autistic participants should find it more difficult to draw on explicit strategies to scaffold implicit learning because the majority of the evidence clearly demonstrates difficulties in explicit serial order memory in autism [Desaunay et al., 2020]. Evidence for the possible contribution of explicit learning in the Bunny Task stems from the finding that participants in Experiment 1 who demonstrated awareness of the repeating bunny sequence also showed the most marked repetition learning effects. Similarly, previous studies have shown that awareness can promote target anticipations in SRT paradigms [Guerard et al., 2011], and Zwart et al. [2017] have provided neural evidence to suggest that autistic participants rely more heavily on explicit rather than implicit learning processes during serial reaction time tasks.

If the Bunny Task is indeed sensitive to explicit rather than implicit serial order learning processes, it would resolve the seeming inconsistency between the current findings and the previous literature. Rather than contradicting evidence of preserved implicit serial order learning, the observations here might simply extend evidence of difficulties in explicit serial order learning to situations where learning occurs passively rather than through explicit instructions. Moreover, the relatively passive nature of the Bunny Task might help explain why group differences were surprisingly marked. Other

tasks that probe explicit serial order memory, such as digit-span or NWR tasks, provide clear instructions for participants to try and remember and then reproduce sequences of stimuli. In the Bunny Task, on the other hand, participants would need to develop an awareness of the repetitions based on passive inferential observation, and this may prove particularly difficult. Future studies could shed further light on these possibilities by employing manipulations that are known to affect the likelihood that participants draw on explicit knowledge in serial reaction time tasks, such as manipulations of the intervals between stimuli [Destrebecqz & Cleeremans, 2001] or the length and complexity of the repeating sequence [Reber & Squire, 1994]. Such manipulations have already been applied to standard manual serial reaction time tasks in studies of autism [e.g., Travers, Klinger, Mussey, & Klinger, 2010] and could easily be extended to the current paradigm, with due consideration of overall task duration (see above). Systematic manipulations of the task across several studies could shed important light not only on the role of serial order learning in relation to language development in autism, but more generally also on the functional integrity of, and interactions between, explicit and implicit learning processes.

Considering the passive demands of the Bunny Task on serial order learning, it should be a particularly useful analogue for the demands of language acquisition, which also primarily occurs passively. Yet, our data did not reveal any correlations between the indices of repetition learning on the Bunny Task and the participants' explicit phonological serial order memory or verbal IQ. The functional consequences of the repetition learning difficulties that are evident on the Bunny task in relation to language development remain therefore somewhat unclear. Interestingly, we did find that learning indices in Experiment 2 were correlated with SCQ scores, indicating that children with the most significant social-communication difficulties demonstrated the most marked repetition learning difficulties. Interpretation of this finding, however, is complicated by the fact that it stands in contrast to a recent finding by Zwart, Vissers, and Maes [2018] who found a positive correlation between learning on an SRT task designed to allow for use of explicit learning strategies and the adult self-report form of the Social Responsiveness Scale [SRS; Constantino & Gruber, 2012] across 35 autistic adults. Specifically, adults who were better at learning a deterministic sequence also reported the greatest social communication difficulties. Moreover, Travers et al. [2010] found no correlation between learning on an SRT task and parent-reported social communication difficulties among a group of autistic and typically developing adolescents. Several authors have examined associations between serial order learning and social-communication difficulties because of the assumption

that serial order learning is important for navigating some of the complex and temporally structured nuances of social interaction. To what extent one can expect to find correlations between broad measures of social-communication behaviors such as the SRS or SCQ and very abstract tasks assessing serial order learning, however, is questionable and the inconsistencies across studies to date may be a reflection of measurement issues. Future studies could address this issue by designing SRT tasks that have greater ecological validity for the potential social-communicative functions of serial order learning.

Before concluding, it is important to acknowledge some important limitations of the current experiments. First, it was unfortunately not possible to acquire a more comprehensive set of language measures to better characterize the verbal abilities (and wider functional profiles) of all children in Experiment 2. Together with a relatively low rate of returns of the SCQ for children in the UK sample, this makes it difficult to fully understand the functional consequences (or correlates) of repetition learning difficulties in ASD. On the other hand, the primary focus for this research was to develop the Bunny Task in the hope that it will prove suitable for assessing serial order learning across heterogeneous groups of participants. In this respect, the current observations hopefully pave the way for future studies to scrutinize the role of serial order learning processes in the heterogeneous language and functional profiles across the autism spectrum. Another potential limitation of the Bunny Task is that it is not entirely clear to what extent it can be considered a reliable and valid measure of serial order learning processes. On the one hand, the group-level data provide robust evidence of serial order learning across the TD participant groups. However, at the level of individual participants, repetition learning is not consistently demonstrated, particularly in Experiment 1. This begs the question of why certain participants demonstrate the phenomenon, while others do not. In one sense this question could be considered trivial because few experimental paradigms elicit phenomena that can be replicated consistently across all participants, particularly when relying on reaction time measures. Such individual differences do not undermine inferences that can be drawn about underlying processes from group-level data. However, when individual differences become the main focus of interest, it is important to know what they represent. Specifically, to make sense of correlations (or lack thereof) between indices of learning demonstrated by the Bunny Task and measures of language function, it is important to know whether individual differences in indices of learning truly reflect underlying variability in learning processes or also other processes (e.g., sustained attention, attention shifting, cognitive flexibility, etc.). Based on the current study alone, we cannot be confident about the processes that might contribute to learning on the Bunny Task and

therefore the conclusions we draw above regarding group differences remain tentative. Future studies will be able to address this shortcoming by scrutinizing the construct validity of the Bunny task against other tasks that are known to probe key processes of interest (e.g., manual SRT tasks).

To conclude, the Bunny Task has the potential to provide important insights into serial order learning processes across the autism spectrum that may play an important role in the heterogeneity of language development. Across two experiments, autistic adults and children demonstrated a marked absence of repetition learning effects on this task, which contrasts with findings of preserved implicit serial order learning in autism but is in line with observations of difficulties on explicit serial order memory tasks. This distinction has attracted considerable interest about the frequently co-occurring language impairments in autism. Along with others who have begun to explore eye-tracking methods as a means to increase our understanding of autistic individuals with complex needs [see Tager-Flusberg et al., 2017] we hope that the current observations will encourage future research on those autistic individuals who remain grossly underrepresented in the literature.

Acknowledgments

Ethics Statement

All procedures were approved by the Psychology Research Ethics Committee at City, University of London, in line with the standards set out by the British Psychological Society and declaration of Helsinki.

References

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders—DSM-IV-TR (4th ed., text revision). Washington, DC: Author.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The special needs and autism project (SNAP). Lancet, 368, 210–215.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism spectrum quotient: Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31(1), 5–17.

Bigham, S., Boucher, J., Mayes, A., & Anns, S. (2010). Assessing recollection and familiarity in autistic spectrum disorders: Methods and findings. Journal of Autism and Developmental Disorders, 40, 878–889.

Boucher, J. (2012). Research review: Structural language in autistic spectrum disorder—Characteristics and causes. Journal of Child Psychology and Psychiatry, 53, 219–233.

Boucher, J., & Anns, S. (2018). Memory, learning and language in autism spectrum disorder. Autism and Developmental Language Impairments, 3, 1–13.

Boucher, J., Mayes, A., & Bigham, S. (2012). Memory in autistic spectrum disorder. Psychological Bulletin, 138, 458–496.

Bowler, D. M., Poirier, M., Martin, J. S., & Gaigg, S. B. (2016). Nonverbal short-term serial order memory in autism spectrum disorder. Journal of Abnormal Psychology, 125, 886–893.

Charman, T., Pickes, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: Data from the special needs and autism project (SNAP). Psychological Medicine, 41, 619–627.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). International Journal of Language and Communication Disorders, 43, 1–40.

Coderre, E. L., Chernenok, M., O'Grady, J., Bosley, L., Gordon, B., & LeDoux, K. (2019). Implicit measures of receptive vocabulary knowledge in individuals with level 3 autism. Cognition, Behaviour and Neurology, 32, 95–119.

Constantino, J., & Gruber, C. (2012). The social responsiveness scale (SRS) (2nd ed.). Los Angeles, CA: Western Psychological Services.

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. Frontiers in Psychology, 6, 1090.

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial Hebb repetition effect. Memory and Cognition, 34, 1720–1729.

Desaunay, P., Briant, A. R., Bowler, D. M., Ring, M., Gérardin, P., Baleyte, J.-M., … Parienti, J.-J. (2020). Memory in autism spectrum disorder: A meta-analysis of experimental studies. Psychological Bulletin, 146, 377–410.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. Psychonomic Bulletin & Review, 8, 343–350.

Elliot, C. D. (2007). Differential ability scales (2nd ed.). San Antonio, TX: Harcourt Assessment.

Falck-Ytter, T., Bölte, S., & Gredebäck, G. (2013). Eye tracking in early autism research. Journal of Neurodevelopmental Disorders, 5, 28.

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, 897–910.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. Applied Cognitive Psychology, 13, 65–77.

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. Journal of Speech, Language and Hearing Research, 50, 177–195.

Guerard, K., Saint-Aubin, J., Boucher, P., & Tremblay, S. (2011). The role of awareness in anticipation and recall performance in the Hebb repetition paradigm: Implications for sequence learning. Memory & Cognition, 39, 1012–1022.

Hermelin, B., & O'Connor, N. (1970). Psychological experiments with autistic children. Oxford, England: Pergamon Press.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. Psychological Bulletin, 140, 339–373.

Hus, V., & Lord, C. (2014). The autism diagnostic observation schedule, module 4: Revised algorithm and standardized severity scores. Journal of Autism and Developmental Disorders, 44, 1996–2012.

Johnson, A. J., Dygacz, A., & Miles, C. (2017). Hebb repetition effects for non-verbal visual sequences: Determinants of sequence acquisition. Memory, 25, 1279–1293.

Johnson, B. P., Lum, J. A. G., Rinehart, N. J., & Fielding, J. (2016). Ocular motor disturbances in autism spectrum disorders: Systematic review and comprehensive meta-analysis. Neuroscience and Biobehavioral Reviews, 69, 260–279.

Lai, W.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educational Research Review, 10, 90–115.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., Dilavore, P. C., … Rutter, M. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders, 30, 205–223.

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, 659–685.

Lum, J. A. G., Conti-Ramsden, G., Morgan, A. G., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. Cortex, 51, 1–10.

Mosse, E. K., & Jarrold, C. (2008). Short article: Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. Quarterly Journal of Experimental Psychology, 61, 505–514. http://dx.doi.org/10.1080/17470210701680779.

Mosse, E. K., & Jarrold, C. (2010). Searching for the Hebb effect in Down syndrome: Evidence for a dissociation between verbal short-term memory and domain-general learning of serial order. Journal of Intellectual Disability Research, 54, 295–307.

Mostofsky, S. H., Goldberg, M. C., Landa, R. J., & Denckla, M. B. (2000). Evidence for a deficit in procedural learning in

children and adolescents with autism: Implications for cerebellar contribution. Journal of the International Neuropsychological Society, 6, 752–759.

Nadig, A., & Mulligan, A. (2017). Intact non-word repetition and similar error patterns in language-matched children with autism spectrum disorders: A pilot study. Journal of Communication Disorders, 66, 13–21. http://dx.doi.org/10.1016/j.jcomdis.2017.03.003.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. Cognitive Psychology, 19, 1–32.

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. G. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. Frontiers in Psychology, 7, 1245.

O'Brien, G., & Pearson, J. (2004). Autism and learning disability. Autism, 8, 125–140.

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. Philosophical Transactions of the Royal Society, B, 364, 3737–3753.

Parmentier, F. B. R., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. Journal of Experimental Psychology: Learning, Memory & Cognition, 31, 412–427.

Poirier, M., Martin, J., Gaigg, S. B., & Bowler, D. M. (2011). Short-term memory in autism spectrum disorder. Journal of Abnormal Psychology, 120, 247–252.

Raven, J. C. (1956). Standard progressive matrices: Sets A, B, C, D, E. London: Lewis.

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. Learning and Memory, 2, 217–229.

Russell, G., Mandy, W., Elliott, D., White, R., Pittwood, T., & Ford, T. (2019). Selection bias on intellectual ability in autism research: A cross-sectional review and meta-analysis. Molecular Autism, 10, 9.

Rutter, M., Bailey, A., & Lord, C. (2003). The social communication questionnaire. Los Angeles, CA: Western Psychological Services.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the autism-spectrum quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. Molecular Autism, 6, 2.

Skwerer, D. P., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. Autism, 20, 591–604.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. Autism Research, 6, 468–478.

Tager-Flusberg, H., Skwerer, D. P., Joseph, R. M., Brukilacchio, B., Decker, J., Eggleston, B., … Yoder, A. (2017). Conducting research with minimally verbal participants with autism spectrum disorder. Autism, 21, 852–861.

Travers, B. G., Klinger, M. R., Mussey, J. L., & Klinger, L. G. (2010). Motor-linked implicit learning in persons with autism spectrum disorders. Autism Research, 3, 68–77.

Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-spatial information: Evidence from eye movement. Psychonomic Bulletin & Review, 13, 452–457.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. Cognition, 92, 231–270.

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. Neuroscience and Biobehavioural Reviews, 51, 205–222.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). Comprehensive test of phonological processing (CTOPP). Austin, TX: PRO-ED.

Wechsler, D. (1991). The Wechsler intelligence scale for children (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). WAIS-III^{UK} administration and scoring manual. San Antonio, TX: Psychological Corporation.

Whitehouse, A. J. O., Barry, J. G., & Bishop, D. V. M. (2008). Further defining the language impairment of autism: Is there a specific language impairment subtype? Journal of Communication Disorders, 41, 319–336.

Williams, D., Botting, N., & Boucher, J. (2008). Language in autism and specific language impairment: Where are the links? Psychological Bulletin, 134, 944–963.

Williams, D., Payne, H., & Marshall, C. (2013). Non-word repetition impairment in autism and specific language impairment: Evidence for distinct underlying cognitive causes. Journal of Autism and Developmental Disorders, 43, 404–417.

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. Journal of Autism and Developmental Disorders, 35, 331–335.

Zwart, F. S., Vissers, C. T. W. M., Kessles, R. P. C., & Maes, J. H. R. (2018). Implicit learning seems to come naturally for children with autism, but not for children with specific language impairment: Evidence from behavioural and ERP data. Autism Research, 11, 1050–1061.

Zwart, F. S., Vissers, C. T. W. M., & Maes, J. H. R. (2018). The association between sequence learning in the serial reaction time task and social impairments in autism. Journal of Autism and Developmental Disorders, 48, 2692–2700.

Zwart, F. S., Vissers, C. T. W. M., van der Meij, R., Kessles, R. P. C., & Maes, J. H. R. (2017). Autism: Too eager to learn? Event related potential findings of increased dependency on intentional learning in a serial reaction time task. Autism Research, 10, 1533–1543.