A physiological marker of recognition memory in adults with autism spectrum disorder? – The pupil Old/New effect

Melanie Ring[1,2], Dermot M. Bowler[2], Sebastian B. Gaigg[2]

[1]Department of Child and Adolescent Psychiatry, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[2]Autism Research Group, Department of Psychology, City, University of London, London, UK

Author note

Melanie Ring is now based at the Department of Child and Adolescent Psychiatry, University Hospital Carl Gustav Carus, Technische Universität Dresden, Fetscherstraße 74, 01307 Dresden, Germany. She was previously based at Autism Research Group, Department of Psychology, City, University of London, Northampton Square, London, EC1V 0HB, United Kingdom, where this research has taken place. Dermot M. Bowler, Autism Research Group, Department of Psychology, City, University of London; Sebastian B. Gaigg, Autism Research Group, Department of Psychology, City, University of London.

This research was presented at the International Meeting for Autism Research 2015 in Salt Lake City, Utah, USA and 2016 in Baltimore, Maryland, USA. It is part of an unpublished doctoral thesis by Melanie Ring entitled "Relational memory processes in adults with autism spectrum disorder", which had been submitted to City, University of London for examination. It is available online through the website of the University library.

Correspondence concerning this article should be addressed to Melanie Ring, Autism Research Group, Department of Psychology, City, University of London, Northampton Square, London, EC1V 0HB, United Kingdom. Email: melanie.ring.1@city.ac.uk or +44(0)20 7040 8380.

Abstract

This study investigated the pupil Old/New effect in individuals with Autism Spectrum Disorder (ASD) and typical development (TD). Participants studied verbal and visual meaningful and meaningless materials in black and white on a computer screen. Pupil sizes were measured while participants performed a Remember (episodic memory with context) /Know (semantic memory, no context) recognition memory test. ASD compared to TD individuals showed significantly reduced recognition rates for all materials. Both groups showed better memory for visual compared to verbal (picture superiority effect) and meaningful compared to meaningless materials. A pupil size ratio (pupil size for test item divided by baseline) for old (studied) and new (unstudied) materials indicated larger pupils for old compared to new materials only for the TD but not the ASD group. Pupil size in response to old versus new items was positively related to recognition accuracy, confirming that the pupil Old/New effect reflects a memory phenomenon in the ASD group. In addition, this study suggests an involvement of the noradrenergic neurotransmitter system in the abnormal hippocampal functioning in ASD. Implications of these findings as well as their underlying neurophysiology will be discussed in relation to current theories of memory in ASD.

Keywords: pupil Old/New effect, episodic/semantic memory, Remember/Know recognition procedure, picture superiority effect, Autism Spectrum Disorder

Lay summary

Most measures of memory in Autism Spectrum Disorder (ASD) depend on verbal answers. In addition to these verbal answers, this study measured the size of the participants' pupil in response to studied and unfamiliar materials revealing memory difficulties in ASD. Measuring pupil size works nonverbally, outside of conscious awareness and forms the basis of studies on less verbal persons with ASD. Mechanisms and brain regions underlying memory differences in ASD are discussed.

Introduction

Memory impairments play a significant role in the functional impairments of many psychiatric illnesses (e.g. depression – Rock et al., 2014; schizophrenia – Aleman et al., 1999) and in the course of healthy and pathological aging (Koen & Yonelinas, 2014). The need for understanding these impairments will exponentially increase as the population ages and memory impairments will become more prevalent. Autism Spectrum Disorder (ASD), which is a disorder defined by difficulties in social interaction and communication as well as restricted and repetitive behaviours (American Psychiatric Association, 2013) is one of the conditions where memory impairments are relatively well documented (Boucher & Bowler, 2008) but still fairly poorly understood. The current study tackles two gaps in the memory literature on ASD. First, it systematically compares long-term episodic and semantic memory for verbal and visual and meaningful and meaningless materials in a large group of high-functioning autistic adults with a broad age-range to see if previous findings transfer to other materials and age groups. Second, it investigates the pupil Old/New effect as a potential physiological marker of memory in ASD to make a first step in the direction of finding measures for memory that would ultimately be suitable for a broad population of individuals with ASD including very young individuals and older adults and persons with intellectual and/or language impairment. The rationale for these aims is provided in the following.

There is a broad consensus that individuals with ASD show specific difficulties with *episodic memory* (Bowler, Gaigg & Lind, 2011) – a memory for personally experienced events including information about time and place and involving a subjective sense of mental time-travel (Tulving, 2002). *Semantic memory* – a memory for timeless facts – is relatively intact in ASD (e.g. Bowler, Gardiner & Gaigg, 2007; Crane & Goddard, 2008; Gaigg, Bowler & Gardiner, 2014). Episodic and semantic memory can be compared directly using the so-called Remember/Know (R/K) recognition procedure (Tulving, 1985), which asks participants whether they *Remember* (R - episodic memory) or *Know* (K - semantic memory) that they have encountered particular information in the past. ASD individuals show specific quantitative impairments in Remembering but not Knowing compared to typically developing (TD) individuals (Bowler, Gardiner & Grice, 2000a; Bowler, Gardiner, Grice & Saavalainen, 2000b; Bowler et al., 2007; Cooper, Plaisted-Grant, Baron-Cohen & Simons, 2017a; Gaigg, Bowler, Ecker, Calvo-Merino & Murphy, 2015; Souchay, Wojcik, Williams, Crathern & Clarke, 2013; Tanweer, Rathbone & Souchay, 2010). When looking at this literature, it is, however, important to note, that the majority of R/K studies in ASD has utilised verbal materials (12 out of 15; e.g. Bowler et al., 2000a&b, 2007; Gaigg et al., 2015), one has used pictures (Souchay et al., 2013), one autobiographical memories (Tanweer et al., 2010), and one non-meaningful kaleidoscope images (Massand, 2011 Experiment 4). No systematic investigations comparing R and K responses have been carried out that compare across these types of materials directly in ASD. However, known language atypicalities (delays or severe difficulties in language development - Baird et al., 2006; Bennett et al., 2008; Boucher, 2012; Loucas et al., 2008; Tager-Flusberg & Joseph, 2003) together with superior perceptual skills may give ASD participants an advantage if pictures were used as materials (Mottron & Burack, 2001; Mottron et al., 2006). Similarly, it is important to consider that ASD individuals have repeatedly been shown to have difficulties to use meaning inherent in study materials (e.g., categorical, semantic, or syntactic information; Bowler et al., 2008a; Frith, 1970a & b; Fyffe & Prior, 1978; Gaigg et al., 2008; Hermelin & O'Connor, 1967; Minshew & Goldstein, 1993; Minshew et al., 1992; Tager-Flusberg, 1991). On the other hand, Ameli, Courchesne, Lincoln, Kaufman and Grillon (1988) found a memory advantage for meaningful materials also in ASD individuals, when comparing memory for meaningful and meaningless visual materials. Therefore, the factor of meaning is important to consider when systematically comparing the influence of material type on memory in ASD.

So far, previous research shows that qualitatively episodic and semantic memories seem to be comparable between groups in that manipulations that differentially affect Remembering and Knowing do so to the same extent in TD and ASD participants (Bowler et al., 2007). Furthermore ASD and TD individuals provide similar justifications for reporting R and K experiences (Bowler et al., 2000a; Souchay et al., 2013; Tanweer et al., 2010), however, the quality of R justifications has not been compared previously. One possibility is that ASD individuals focus on details included in the immediate study context to justify their Remember responses because of reduced generalisation across different contexts (Plaisted, O'Riordan & Baron-Cohen, 1998). With less information to base their judgements on, this could be related to difficulties distinguishing old and new items and increased false alarm rates (e.g. Gardiner, Bowler & Grice, 2003). Qualitative differences are also supported by recent neurophysiological evidence suggesting that both episodic and semantic memory are associated with atypical patterns of neural activity in ASD (Gaigg, Bowler, Ecker, Calvo-Merino, & Murphy, 2015; Massand, Bowler, Mottron, Hosein & Jemel, 2013; Massand & Bowler, 2015; Cooper, et al., 2017b). These studies, however, also show that covert physiological indices exist that can index the memory atypicalities that characterise the disorder. This is important, because such indices will be critical for understanding what role memory difficulties play in the early development of ASD, when infants are not yet able to cope with the demands of more explicit memory tests. Neural measures, however, are difficult to obtain from infants and therefore the current study asks whether pupil dilation might also serve as a physiological marker of memory difficulties in ASD. Pupil size responses can be measured with very minimal demands on participants, they are a reflex that exists from birth and that operates independently of conscious awareness (Gomes, Montaldi & Mayes, 2015; Heaver & Hutton, 2011; Laeng, Sirois & Gredebäck, 2012). Similarly to ERPs, pupil responses have a good temporal sensitivity (Hartmann & Fischer, 2014), but in contrast to ERPs, they are relatively easy and cheap to record (Laeng et al., 2012).

Small correlations between pupil dilation and ERPs indicate that the two measures assess different underlying processes (Steinhauer & Hakerem, 1992). Because of these advantages, pupillometry seems in principle a good measure to test a wide range of individuals including less verbal ones. Traditionally, pupil dilation was seen as an indicator of cognitive load, for example, in working memory (e.g., Piquado, Isaacowitz & Wingfield, 2010) or visual search tasks (Porter, Troscianko & Gilchrist, 2007), where greater pupil dilation indicates higher cognitive load and an overload is demonstrated by a decrease in pupil size, possibly resulting from task disengagement. Regarding memory, TD individuals have reliably been reported to show larger pupils in response to previously studied (old) compared to unstudied (new) items in a phenomenon known as the pupil Old/New effect. A series of studies indicated the pupil Old/New effect to be influenced by memory strength (Otero, Weekes & Hutton, 2011; Papesh, Goldinger & Hout, 2012), emotion (Võ et al., 2008), and the degree to which encoding and retrieval conditions matched (Papesh et al., 2012). It was found to be universal across different materials (Otero et al., 2011), and pupil size at encoding and retrieval distinguished between later correctly and falsely remembered materials (Montefinese, Ambrosini, Fairfield & Mammarella, 2013; Otero et al., 2011; Papesh et al., 2012). Recently, pupil dilation at encoding has been related to successful free recall of items emphasizing its potential as a biomarker of memory (Kucewicz et al., 2018). In addition, investigations in amnesia show the potential of pupil responses to indicate memory abnormalities in that amnesic individuals show larger pupils in response to new as opposed to previously studied materials (i.e. a novelty response - Laeng et al., 2012), indicating the potential of pupil size measurements to reveal memory abnormalities in patient groups.

Following the literature reviewed above, our aims for the current study were the following. First, it was of interest to examine episodic and semantic memory across various types of stimuli in ASD and TD participants across a large age-range to investigate whether previous results of reduced Remembering and intact Knowing in ASD can be replicated and generalised across age and different verbal and visual meaningful and meaningless materials. By investigating remember justifications, it was of relevance to inspect on what criteria participants base their recognition judgements and whether they may differ between groups. It was predicted that ASD individuals would rely less on information from outside the immediate study context due to less generalisation. Second, we aimed to establish whether pupil dilation Old/New effects might provide a covert index of memory impairments in ASD that could be used in future studies of much younger and/or less able participants. We predicted the well-established pupil Old/New effect for TD individuals. Based on ERP Old/New studies in ASD (e.g. Massand & Bowler, 2015), it was expected that the pupil Old/New effect would either be diminished or enhanced in ASD as compared to TD individuals.

Methods

Participants

Of the 64 participants (32 in each group) tested, pupillometry data were not available for five ASD (four men, M_{age} = 49.91 years, age range: 32-65, $M_{\text{VIQ/VCI}}$ = 113, $M_{\text{PIQ/PRI}}$ = 101, M_{FIQ} = 108) and two TD individuals (two men, M_{age} = 41.10 years, age range: 36-46, $M_{\text{VIQ/VCI}}$ = 96, $M_{\text{PIQ/PRI}}$ = 94, M_{FIQ} = 95), who did not differ significantly from the rest of the sample. The final sample consisted of twenty-seven ASD adults (23 men, M_{age} = 42.31 years, age range: 27-64 years) and 30 TD individuals (23 men, M_{age} = 43.98 years, age range: 22-65 years), who were individually matched on Verbal Intelligence Quotient (VIQ), Performance IQ (PIQ) and Full-scale IQ (FIQ) as measured by the third or fourth edition of the Wechsler Adult Intelligence Scale (WAIS-III$^{\text{UK}}$ or WAIS-IV$^{\text{UK}}$; The Psychological Corporation, 2000; 2008; see Table 1). In addition, groups were closely matched on gender, X^2 = 0.66, p = .42, and chronological age. Participants were recruited through a database of individuals with whom the

Autism Research Group is in regular contact and in addition through newspaper advertisements, flyers and word of mouth. All participants were native English speakers. All ASD individuals had received a clinical diagnosis according to DSM-IV-TR criteria (American Psychiatric Association, 2000) prior study and 21 of these individuals were available to take part in the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1989) administered by researchers trained to research reliability standards on this instrument. Five of these individuals scored just below the total cut off score of 7 for ASD. They were nevertheless included in the study since records confirmed that they all had a clinical diagnosis of an ASD which was our main inclusion criterion. TD individuals were included if they reported that they did not take psychotropic medication or that they had no personal or family history of a psychological or neurodevelopmental disorder. All participants were reimbursed for their time and travel expenses according to standard university fees. This study was approved by the ethics committee of the Psychology Department of City, University of London and the procedures used in this study adhered to the ethical guidelines set out by the British Psychological Society and were in accordance with the provisions of the World Medical Association Declaration of Helsinki.

[Insert Table 1 here]

Materials

The materials used in this study included words, pictures, shapes and non-words (see Figure 1 for examples). Pictures were selected from Snodgrass and Vanderwart (1980) and the words comprised the relevant labels for these pictures. Shape stimuli were provided by Haenschel et al. (2007) and non-words were selected from Gathercole, Willis, Emslie and Baddeley (1991). Two of the original non-words turned out to be meaningful words and were therefore replaced. The final set of materials was selected based on a number of pilot studies that served to ensure that all pictures could be named clearly, that shapes were not too difficult to remember and that the difficulty level for all materials avoided ceiling or floor effects. The final set of materials included four lists with 10 items each for pictures and words. Pictures and words were selected as the same lexical item and their presentation was counterbalanced in a way that participants never saw the same item as picture and word during study or test and that each item was presented equally often either as a word or picture and either as a to-be-remembered target or lure item. The four lists were matched for letter number, name agreement, image agreement, familiarity and complexity ratings and word frequency (Kucera and Francis; see Table 2).

[Insert Table 2 here]

In addition, we included two lists of 10 two and three syllable non-words and two lists of 10 shapes that proved to be easiest to name in a pilot study including 120 items. The two lists were counterbalanced across participants so that each item equally often served as either a to-be-remembered or lure item in the memory test.

[Insert Figure 1 here]

To avoid pupil size estimation bias resulting from eye movements (Brisson et al., 2013), the above described materials were presented in the centre of the screen and images were large enough to recognise but small enough to see without the need for eye movements. At the participant’s viewing distance of approximately 60cm the shapes measured 2.85 x 2.86 degrees, pictures had the size of 4.77 x 3.82 degrees, and words and non-words measured 5.72 x 1.91

degrees of visual angle. Avoiding systematic changes in pupil size because of the pupil light reflex, all materials were presented on the same grey background and images were black and white to ensure similar luminance within and across conditions. Old (studied) and new (unstudied) sets of materials were counterbalanced across participants to control for any systematic differences in luminance between sets. In addition, the items were presented in blocks of the same material type, i.e., shapes, pictures, words, and non-words were never intermixed within one block. Luminance measurements for shapes and pictures were taken across the whole image using a Konica Monolta LS-100 luminance meter. Although meaningful pictures (M = 159.28 cd/m^2, SD = 11.65) were significantly brighter, t(43.10) = 22.16, p < .0001, Cohen's d = 4.30, 95 % CI(3.36, 5.24), than meaningless shapes (M = 117.36 cd/m^2, SD = 1.92)[1], the different sets of shapes, t(18) = 0.15, p = .89, Cohen`s d = 0.06, 95 % CI(-0.81, 0.94), and pictures, F(3,36) = 1.32, p = .28, η^2 = .10, were well matched in terms of luminance (see Table 3).

[Insert Table 3 here]

Procedure

Remember/Know recognition test

Materials were presented on a Tobii TX300 remote eye-tracking screen. During study, participants were asked to memorise a set of words, pictures, shapes and non-words that was presented in blocks of 10 items each according to their material type. The presentation order of the blocks was counterbalanced across participants using a Latin Square. This created 16 different versions of the task which were presented so that matched pairs of participants (an

[1] It is unlikely that the luminance difference between pictures and shapes confounded the results because pupil size data were collapsed across the different material types and only analysed that way. In addition, this difference in luminance was the same for the different lists and versions of the experiment and, therefore, counterbalanced across participants.

ASD and a TD participant with similar IQ) saw identical orders of the materials[2]. The order of items within blocks was randomised. Each item was presented for 2 seconds. Between two items a blank screen was presented for 1 second (see Figure 2). The test followed straight after study. In the test, participants saw 80 items in 4 blocks of the different materials with 20 items each. Half of the items were the items from study. The order of presentation of the blocks was the same as at study so that there was the same length of time between study and test phase for all the materials. Participants were asked to indicate which items they had seen previously by pressing the appropriate key on the keyboard. In addition, for items they reported having seen before, they were asked if they could remember clearly the context or time of studying the item in addition to the item, i.e. any information about the actual study episode (R/type 'a' memory)[3] or if they simply knew the item was presented without remembering any additional contextual information (K/type 'b' memory; see Figure 2 for an illustration of the procedure of the task). Before the experiment, examples of these types of memory were given and during the recognition test, participants were instructed to justify their responses. At test, participants first saw a fixation cross for 250 milliseconds (baseline condition for pupil size measurements), which was followed by the presentation of the actual item for 1.75 seconds. After that participants saw a 'Yes - No' and in case they pressed Yes, a 'Type a – Type b' answer screen until they gave their response by pressing the appropriate key on the keyboard (see Figure 2).

[Insert Figure 2 here]

Episodic memory in Remember judgements

[2] Because pupil data were not available for seven participants (five ASD and two TD), three TD individuals did not have a matched ASD individual taking part in the same version of the task and were therefore group-matched on IQ, gender and age.

[3] Typical responses for a remember response could be: "This was the first item in the list.", "This reminded me of my last holiday.", "This shape looked a bit like an island and therefore I remembered it.". For further examples see section*: Episodic memory in Remember judgements*

All participants provided detailed justifications for all R responses, and it was of interest to inspect the quality of R justifications. Therefore justifications for R responses were tape-recorded, transcribed and classified according to the kind of statements that were produced. One logical way of coding was to inspect the number of associations participants formed with information from the immediate study episode and information relating to general knowledge or personal experiences that had not directly been part of the study. Therefore, statements were categorised into two groups - things that happened *within* the actual study episode (item characteristics, e.g., “I remember this because it was just one sock.”, or study episode, e.g., “I pictured the word in my head.”), and information from *outside* the study episode that the participant had related to the actual items (semantic knowledge, e.g., “I remembered the apple because it is a fruit.”, or personal experiences, e.g., “I had a sandwich for lunch.”). A second independent rater, who was blind to the predictions and groups, and who had been trained on these criteria, scored the transcripts of eight randomly selected participants (four from each group). The strength of the inter-rater agreement between the scorings of the first author and the second rater, calculated with Cohen’s kappa, was very good, $\kappa = .827$, $p < .0001$, showing that the aforementioned scheme can be coded reliably. In addition, there were no significant differences between groups in terms of the length of the audio recordings ($M_{ASD} = 10.09$, $SD_{ASD} = 2.54$; $M_{TD} = 10.07$, $SD_{TD} = 1.91$), $t(55) = 0.04$, $p = .97$, Cohen’s $d = 0.01$.

Pupillometry

After a standard five-point calibration procedure, pupil diameter was recorded throughout the task with a Tobii TX300 eye-tracker with a sampling rate of 120 Hz. Customised Matlab routines were used to remove artefacts, linearly interpolate blinks, and to extract the data. For linear interpolation, the five samples before a blink and the five samples after a blink were averaged and linearly interpolated so that missing values were incremented ending up with a straight line connecting values before and after the blink. Applying a low-pass Butterworth filter with a cut-off frequency of 1 Hz, high-frequency noise in the data (e.g. caused by partial blinks) was removed. The data for each eye were interpolated and filtered before averaging them across the two eyes. There was no significant difference between groups in the number of excluded trials (absolute number: $M_{ASD} = 4.26$, $SD_{ASD} = 1.51$, $M_{TD} = 4.00$, $SD_{TD} = 1.53$, percentage: $M_{ASD} = 0.05$, $SD_{ASD} = 0.02$, $M_{TD} = 0.05$, $SD_{TD} = 0.02$, $p = .52$, Cohen's $d = 0.17$, 95 % CI(-0.35, 0.69)). No participants had to be excluded for too many partial blinks or bad data quality at this point. A pupil size ratio was then calculated (Heaver & Hutton, 2011) to control for natural pupil size fluctuation and differences in pupil size between participants at baseline. For this, the maximum pupil size during item presentation (i.e., the task-evoked pupillary reflex) was divided by the maximum pupil size at baseline (i.e., pupil size at the presentation of the fixation cross before item presentation). Ignoring data for the first test trial to reduce noise in the data following the change from study to test, the data were averaged across trials separately for studied (old) and unstudied (new) items.

Statistical analysis

The behavioural raw data were scored in terms of hit rates (percentage of yes responses to studied items), false alarm (FA) rates (percentage of yes responses to lure items) and corrected recognition rates (Hits minus FA). Results were analysed using Chi-Squared tests, independent samples t-tests, repeated measures ANOVAs, bivariate correlations and (multiple) linear regressions. In the case of significant differences, Bonferroni corrected post hoc tests were used. Greenhouse Geisser correction (GGC) was applied when the Sphericity assumption was violated. The level of significance was set to .05. Cohen's *d* and partial eta squared are reported as measures of effect size.

Results

Accuracy

The data for corrected recognition are presented in Table 4 and Figure 3 and they were analysed using a 2 (Group [ASD, TD]) x 2 (Modality [verbal, visual]) x 2 (Meaning [meaningful, meaningless]) x 2 (R/K [Remember, Know]) repeated measures ANOVA. This showed significant main effects of *Group*, $F(1,55) = 15.22$, $p < .0001$, Cohen's $d = 1.04$, 95 % CI(0.47, 1.57), with higher corrected recognition rates for TD compared to ASD individuals, *Modality*, $F(1,55) = 4.46$, $p = .04$, Cohen's $d = 0.35$, 95 % CI(-0.02, 0.72), with higher corrected recognition for visual compared to verbal materials, *Meaning*, $F(1,55) = 19.48$, $p < .0001$, Cohen's $d = 0.65$, 95 % CI(0.27, 1.02), with higher corrected recognition for meaningful compared to meaningless materials, and *R/K*, $F(1,55) = 127.61$, $p < .0001$, Cohen's $d = 2.56$, 95 % CI(2.05, 3.03), with higher corrected recognition for R compared to K responses. The data were further characterised by a number of interactions. First, a two-way interaction of *Group x R/K*, $F(1,55) = 8.65$, $p = .01$, $\eta_p^2 = .14$, confirmed higher corrected R recognition in the TD compared to the ASD group ($p < .0001$, Cohen's $d = 1.01$, 95 % CI(0.44, 1.54)), with no difference in corrected K recognition (see Figure 3). Second, a *Modality x Meaning* interaction, $F(1,55) = 31.91$, $p < .0001$, $\eta_p^2 = .37$, demonstrated higher corrected recognition for meaningful pictures as compared to meaningless shapes ($p < .0001$, Cohen's $d = 1.17$, 95 % CI(0.76, 1.55)) with no effect of meaning on memory for words vs. non-words (see Figure 3).

[Insert Figure 3 here]

Third, a *Modality x R/K* interaction, $F(1,55) = 5.91$, $p = .02$, $\eta_p^2 = .10$, showed higher corrected R responses for visual compared to verbal materials ($p = .03$, Cohen's $d = 0.38$, 95 % CI(0.01, 0.75)) and no difference between these materials for corrected K responses. Finally, a significant three-way interaction of *Modality x Meaning x R/K*, $F(1,55) = 5.17$, $p = .03$, $\eta_p^2 = .09$, further qualified the interactions among these factors. Specifically, the effect of meaning on recognition performance was evident in corrected R responses. Meaningless non-words were better recognised than meaningful words ($p = .04$, Cohen's $d = 0.32$, 95 % CI(-0.05, 0.68)), whereas meaningful pictures were better recognised than meaningless shapes ($p < .0001$, Cohen's $d = 0.66$, 95 % CI(0.28, 1.03)). Corrected K responses did not differ for verbal materials but corrected K responses were higher for meaningful pictures compared to meaningless shapes ($p < .0001$, Cohen's $d = 0.48$, 95 % CI(0.31, 1.07)). The absence of any additional interactions involving the group factor, $F_{max} < 1.11$, $p_{min} > .29$, $\eta_p^2{}_{max} < .03$, indicates that the attenuated levels of R but not K responses in the ASD group are persistent across meaningful and meaningless verbal and visual materials.

[Insert Table 4 here]

False alarms

The data are presented in Table 4. To examine whether the attenuated levels of remembering in the ASD group were the result of a more lenient response criterion specifically for R responses in this group, FAs were analysed using a 2 (Group [ASD, TD]) x 2 (Modality [verbal, visual]) x 2 (Meaning [meaningful, meaningless]) x 2 (R/K [Remember, Know]) repeated measures ANOVA (see Table 2). This showed a marginal main effect of *Group*, $F(1,55) = 3.19$, $p = .08$, Cohen's $d = 0.33$, 95 % CI(-0.04, 0.70), with higher FAs for the ASD compared to the TD group as well as a significant main effect of *Meaning*, $F(1,55) = 4.98$, $p = .03$, Cohen's $d = 0.33$, 95 % CI(-0.04, 0.69), with higher FAs for meaningless compared to meaningful materials. A significant *Modality x Meaning* interaction, $F(1,55) = 6.08$, $p = .02$, $\eta_p^2 = .10$, showed higher FAs for meaningless shapes compared to meaningful pictures ($p = .003$, Cohen's $d = 0.54$, 95 % CI(0.16, 0.91)), but no effect of meaning on verbal materials. No other main effects or interactions were significant, $F_{max} < 1.85$, $p_{min} > .17$, $\eta_p{}^2{}_{max} < .04$, including the interaction between Group x RK, making it unlikely that differences in response criteria are the source of the Group x RK interaction in the corrected recognition rates.

Sensitivity and response bias

Because one could argue that Remember and Know responses are not independent from each other and sensitivity and response bias are confounded in measures such as Hits, FAs, and corrected recognition rates (Stanislaw & Todorov, 1999), further analyses were performed using separate measures for sensitivity and response bias. Since d' assumptions cannot be tested in a yes-no task (Stanislaw & Todorov, 1999), A' (Pollack & Norman, 1964) was used as a measure of sensitivity in the current study, as it is nonparametric and unaffected by response bias (Stanislaw & Todorov, 1999). Response bias was estimated by the nonparametric measure B" (Grier, 1971).

A' Remember data (presented in Table 5) were analysed with a 2 (Group [ASD, TD]) x 2 (Modality [verbal, visual]) x 2 (Meaning [meaningful, meaningless]) repeated measures ANOVA, which showed significant main effects of *Group*, $F(1,55) = 12.52$, $p < .01$, Cohen's $d = 0.93$, 95 % CI(0.38, 1.47), with higher A' rates for the TD compared to the ASD group, and *Meaning*, $F(1,55) = 5.47$, $p < .05$, Cohen's $d = 0.35$, 95 % CI(-0.02, 0.72), with higher A' rates for meaningful compared to meaningless materials. There was also a significant *Modality x Meaning* interaction, $F(1,55) = 12.40$, $p < .01$, $\eta_p^2 = .18$, with higher A' rates for meaningful pictures compared to meaningful words, $p < .0001$, Cohen's $d = 0.69$, 95 % CI(0.31, 1.06), but no difference between meaningless non-words compared to meaningless shapes, $p = .24$, Cohen's $d = 0.22$, 95 % CI(-0.15, 0.58). No other main effects or interactions were significant, $F_{max} < 1.57$, $p_{min} > .21$, $\eta_p{}^2{}_{max} < .03$. Overall these data indicate that ASD individuals had lower sensitivity in their answers, i.e. difficulties in distinguishing old and new items.

[Insert Table 5 here]

B" Remember rates (see Table 5) were analysed with a 2 (Group [ASD, TD]) x 2 (Modality [verbal, visual]) x 2 (Meaning [meaningful, meaningless]) repeated measures ANOVA. There was only a marginal *Group x Modality x Meaning* interaction, $F(1,55) = 3.61$, $p = .06$, $\eta_p{}^2 = .06$, with slightly higher B" Remember rates for meaningful words for TD compared to ASD individuals and meaningful pictures for ASD compared to TD individuals. No other main effects or interactions were significant, $F_{max} < 1.79$, $p_{min} > .18$, $\eta_p{}^2{}_{max} < .04$.

Episodic memory in Remember judgements

The data were analysed using a 2 (Group [ASD, TD]) x 2 (Type of EM statements [inside, outside]) repeated measures ANOVA. The analysis showed significant main effects of *Group*, $F(1,55) = 11.10$, $p = .002$, Cohen's $d = 0.88$, 95 % CI(0.33, 1.42), with a higher number of EM statements for the TD compared to the ASD group, and *Type of EM statements*, $F(1,55) = 7.75$, $p = .007$, Cohen's $d = 0.61$, 95 % CI(0.23, 0.99), with more statements referring to information from outside compared to within the study episode. A marginal *Group x Type of EM statements* interaction, $F(1,55) = 3.94$, $p = .05$, $\eta_p{}^2 = .07$, showed more EM statements from outside the study episode for the TD ($M = 16.53$, $SD = 7.64$) compared to the ASD group ($M = 10.11$, $SD = 7.64$), $p = .002$, Cohen's $d = 0.84$, 95 % CI(0.29, 1.37), but a similar number of EM statements from inside the study episode for both groups (M_{TD} = 9.23, SD_{TD} = 6.12; M_{ASD} = 8.89, SD_{ASD} = 6.12), p = .83, Cohen's d = 0.06, 95 % CI(-0.46, 0.58).

Pupil Old/New effect

The data are presented in Figure 4. They were analysed with a 2 (Group [ASD, TD]) x 2 (Set [Old, New]) repeated measures ANOVA. There was a significant main effect of *Set*, $F(1,55) = 31.98$, $p < .0001$, Cohen's d = 0.26, 95 % CI(-0.11, 0.63), with a larger pupil for old compared to new items (i.e., the expected Old/New effect). There was also a significant *Set x Group* interaction, $F(1,55) = 11.31$, $p = .001$, $\eta_p^2 = .17$, with a significantly larger pupil for old compared to new items for the TD group ($p < .0001$, Cohen's d = 0.42, 95 % CI(-0.10, 0.93)) but a similar pupil size for old and new items for the ASD group (p = .12, Cohen's d = 0.11, 95 % CI(-0.43, 0.64)). The main effect of Group was not significant, $F(1,55) = 0.44$, p = .51, Cohen's d = 0.18, 95 % CI(-0.35, 0.69).

[Insert Figure 4 here]

Relation between pupil size and behavioural data

We investigated these results further by analysing the two groups separately. The analysis showed a significant positive correlation for the ASD group (r = .47, p = .01) indicating that a larger pupil dilation at test in response to old versus new items was related to higher corrected recognition scores. There was no significant relation for the TD group (r = -.12, p =

.53; see Figure 5) [4]. These correlations provide additional evidence that the pupil Old/New effect can serve as an index picking up broad group differences in memory function.

[Insert Figure 5 here]

Exploratory investigation of the effects of age

Since the groups of individuals tested include a large age-range, it was of interest to investigate the effect of age on the memory indices used in this study. First, bivariate correlations were run, which showed no significant correlations between corrected Remember responses (Hits minus false alarms; r = -0.01, *p* = .97), corrected Know responses (Hits minus false alarms; r = 0.13, *p* = .35), an Old-New pupil response ratio (pupil response to Old minus pupil response to New unstudied items; r = 0.20, *p* = .15) and age.

It is possible, however, that group as a third variable may have influenced the relationship between Remember, Know, pupil responses and age (Bewick, Cheek & Ball, 2003), as age may have had a differential effect on both groups. Therefore, multiple linear regression analyses to predict memory performance including Age and a Group x Age interaction term were calculated in a second step.

The Group x Age interaction term explained 18.1 % of the total variance, R^2 = .18, $F(1,55) = 12.15$, $p = .00$, and it significantly predicted corrected Remember responses, β = -.43, 95 % CI(-0.01, 0.00), $p < .01$. Visual inspection of Figure 6 (left) and the regression coefficients showed that age was a better predictor of Remember responses in the TD as opposed to the ASD group. Similarly, a model including Age and the Group x Age interaction

[4] Closer inspection of the data showed 4 outliers in our dataset (1 ASD and 1 TD individual with a very small pupil size, 1 TD individual with a very large pupil size and 1 TD individual with a very low corrected recognition score). Outliers in this case were defined as having scores of more than two standard deviations above or below the group mean. Excluding these outliers provided similar results to the ones provided above (ASD: r = .61, *p* = .001; TD: r = -.25, *p* = .20). Therefore, these individuals were kept in the sample.

term explained 22.2 % of the total variance, $R^2 = .22$, $F(2,54) = 7.68$, $p = .00$, and Age, $\beta = -.43$, 95 % CI(0.00, 0.00), $p < .01$, and Group x Age interaction, $\beta = .27$, 95 % CI(0.00, 0.00), $p < .05$, significantly predicted an Old-New pupil response ratio. Visual inspection of Figure 6 (right) and the regression coefficients showed that age was a better predictor of Old-New pupil response ratio in the TD as opposed to the ASD group.

Neither Age nor a Group x Age interaction term explained any variance in corrected Know responses.

[Insert Figure 6 here]

Discussion

Our main aims of this study were to systematically examine episodic and semantic memory across various types of materials in adults with and without ASD across a large age-range and to establish whether the pupil Old/New effect could serve as a covert physiological index of memory difficulties in this group.

Confirming our first prediction, we found higher R rates for TD compared to ASD individuals suggesting specific difficulties with episodic but not semantic memory in ASD. This finding replicates previous literature (e.g. Bowler et al., 2000a; 2007; Souchay et al., 2013; Gaigg et al., 2015). Our findings were further qualified by significant main effects. Regarding modality, we found that both groups remembered visual materials better than verbal materials, confirming the *picture superiority effect* (Shepard, 1967) in both groups. ASD and TD individuals remembered meaningful materials better than meaningless materials, indicating that both groups found it easier to use meaning that was inherent in the study materials rather than to establish meaning for the materials themselves. The reduction of R responses in the ASD group was persistent across these effects. Since R responses require remembering the items as well as the context of their presentation, these data suggest particular difficulties with the retrieval of relational information in ASD (Bowler et al., 2011). The participants' justifications for R responses lend further support by showing that ASD individuals related studied materials significantly less to information they knew from outside the immediate study episode, such as semantic knowledge or personal experiences. This finding fits with the increased perceptual discrimination account of ASD (Plaisted et al., 1998a) and the notion that ASD individuals find it difficult to transfer information from one context to another and to generalise from experiences. A recent study (Sapey-Triomphe, Sonié, Hénaff, Mattout & Schmitz, 2018) takes these ideas further by showing differences in learning styles. Whereas ASD individuals memorised and recalled each episode separately ('look-up-table strategy'), TD individuals interpolated between exemplars to generalise across experiences ('interpolation strategy'). For the current study this would indicate that ASD individuals showed difficulties in recruiting meaning to support memory while relying on information that was presented in the test, whereas TD individuals imagined possibilities and formed relations to generalised information in their memory. In a more general context, these findings are in line with a recurring theme in ASD research namely the *failure to encode information meaningfully* by Hermelin and O'Connor (1967) who tested the observation that ASD individuals showed a good memory for facts often without an understanding of meaning or without being able to place them into context. The current study expands these earlier findings with verbal material to visual and meaningless materials and indicates that difficulties are not specific to aspects of language but are of a more general nature, hinting at problems with relational processing in ASD (Bowler et al., 2011).

In line with our second prediction and a growing body of literature (Gomes et al., 2015; Heaver & Hutton, 2011; Otero et al., 2011; Papesh et al., 2012), we found the pupil Old/New effect for TD individuals. In contrast, ASD individuals showed similar pupil sizes for old and new items, suggesting that, physiologically, ASD individuals did not distinguish between old and new items. This finding parallels the ERP findings from Massand and Bowler (2015) who observed an absence of some of the prominent ERP Old/New effects known from the TD literature for ASD individuals. Physiologically, pupil responses are controlled by the locus coeruleus, which is related to the hippocampus via noradrenergic transmissions and it, therefore, either inhibits or enhances hippocampal functions, having an important role for long-term memory consolidation and retrieval shortly after study (Sara, 2009). Abnormalities in processes driven by the hippocampus have been reported previously in ASD (e.g. Ring, Derwent, Gaigg and Bowler, 2017), however, the current study adds to this literature by indicating a possible pathway of alternation: reduced or increased transmitter release could lead to an alternation of hippocampal functioning with reduced consolidation and retrieval in ASD ultimately producing weaker memories indicated by the lack of the pupil Old/New effect. This interpretation fits well with the behavioural memory data, showing slightly higher FAs and lower sensitivity in the ASD group also indicating a difficulty in distinguishing studied from new unstudied items. Further support comes from previous research showing higher K rates in R/K recognition tests (Bowler et al., 2000a; 2007) and more intrusion errors in free recall tests (e.g. Bowler et al., 2000b, 2008a; Kamio & Toichi, 2007; Tager-Flusberg, 1991) and work on memory illusions (Gaigg & Bowler, 2009). Taken together, these results suggest some level of confabulation in ASD, which may be related to a problem with metacognition in terms of response monitoring. Difficulties in this area in ASD have been reported previously (Grainger, Williams & Lind, 2014; Wilkinson, Best, Minshew & Strauss, 2010; Wojcik, Moulin & Souchay, 2013).

A lack of a difference in pupil size in response to old vs new items could also show an information overload indicated by a “levelling” of the pupil size leading to task disengagement, poorer encoding and ultimately weaker memories with difficulties in distinguishing old and new items at test. This idea is supported by findings from typically developing older people, whose pupil size did not distinguish between different levels of memory load (Van Gerven et al., 2004). This similarity between findings in ASD and typical older people provides support for the *ageing analogy* of autistic people's memory (Bowler, 2007), which suggests that there are parallels in memory functioning between individuals with ASD and TD older people. Findings from the exploratory regression with age lead in a similar direction. Age had a differential effect on R responses in the two groups, in that a stronger age-related memory difference was found for TD as opposed to ASD individuals. Younger and older ASD individuals performed similarly and at a lower level compared to the TD individuals. This finding is in line with the *safeguard hypothesis* (Geurts & Vissers, 2012), and recent studies reporting reduced effects of age on visual memory in (large) ASD compared to TD samples (Lever & Geurts, 2016; Ring, Gaigg & Bowler, 2016). The current study expands this earlier work to verbal and meaningful and meaningless materials.

Leaving aside questions about the ultimate source of absent pupil Old/New effects in ASD, the fact that this effect discriminated between ASD and TD participants is important because it provides new opportunities for addressing some critical questions concerning the role of memory impairments in the aetiology of ASD. First, pupillary responses can be measured with relative ease from a very young age alongside other eye-tracking measures that have proven fruitful for the examination of the early development of memory (e.g., Richmond & Nelson, 2009). In combination, therefore, eye-tracking indices have great potential for shaping our understanding of how early-emerging memory phenomena might deviate from a typical developmental trajectory in ASD and how such deviance might predict later emerging core and associated features of the ASD clinical phenotype. Similarly, eye-tracking technology has great potential for developing our understanding of memory strengths and weaknesses in the approximately 30% of autistic individuals who remain functionally minimally verbal, and who remain grossly underrepresented in the research literature. Some have argued that memory impairments are among the critical causes of the language difficulties that characterise this group, yet it has proven extremely difficult to develop appropriate experimental methods to address this prediction systematically (see Boucher, Mayes & Bigham, 2012). Significant correlations between the pupil Old/New effect and behavioural memory accuracy in the ASD group suggest that for the ASD individuals that could distinguish between previously studied and new unstudied items, larger corrected recognition was related to a larger difference in pupil dilation towards old vs. new items. Interestingly, this correlation was absent in the TD group. The reason could be a potential ceiling effect in corrected recognition. Most interesting is, however, the fact that the variability in the ASD group modestly predicts variability in their memory which does not happen in the TD group. This finding highlights the potential of the pupil Old/New effect as a diagnostic marker for ASD. It is possible that it differentiates between a non-verbal child with ASD without a pupil Old/New effect and a non-verbal child with another developmental delay that shows a pupil Old/New effect. More research is needed in this direction showing whether the absence of a pupil Old/New effect in ASD can be replicated, which other behavioural memory measures it is related to and whether this is an effect that is specific to ASD or whether it is also present in other clinical populations such as individuals with Schizophrenia who already share other memory characteristics with ASD individuals (e.g. Goldstein, Minshew, Allen & Seaton, 2002).

To conclude, the current study is the first systematic comparison of recognition memory for visual and verbal meaningful and meaningless materials in a large well-matched adult sample of ASD and TD individuals. We reported the picture superiority effect for ASD individuals and showed that memory for meaningful materials was better than for meaningless materials. We replicated previous findings of intact semantic memory but decreased episodic memory in ASD and were able to specify these difficulties even further. In addition, we replicated these difficulties using a physiological measure supporting previous findings of different underlying neurophysiology in ASD. The paradigm used in the current study seems to be a suitable one to evoke pupil size responses and future research should try to establish more precisely the parameters that influence this important marker of memory.

References

Aleman, A., Hijman, R., de Haan, E. H., & Kahn R. S. (1999). Memory impairment in schizophrenia: A meta-analysis. *The American Journal of Psychiatry, 156*(9), 1358-1366.

Ameli, R., Courchesne, E., Lincoln, A., Kaufman, A. S., & Grillon, C. (1988). Visual memory processes in high-functioning individuals with autism. *Journal of Autism and Developmental Disorders, 18*(4), 601-615.

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

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (Fifth ed.).* 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). *The Lancet, 368*(9531), 210-215.

Bennett, T., Szatmari, P., Bryson, S., Volden, J., Zwaigenbaum, L., Vaccarella, L., … Boyle, M. (2008). Differentiating autism and Asperger syndrome on the basis of language delay or impairment. *Journal of Autism and Developmental Disorders, 38*(4), 616-625.

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical Care, 7*(6), 451-459.

Boucher, J. (2012). Putting theory of mind in its place: Psychological explanations of the socio-emotional-communicative impairments in autistic spectrum disorder. *Autism, 16*(3), 226-246.

Boucher, J. & Bowler D. M. (Eds.). (2008). *Memory in autism: theory and evidence.* Cambridge: Cambridge University Press.

Boucher, J., Mayes, A., & Bigham, S. (2012). Memory in autistic spectrum disorder. *Psychological Bulletin, 138*(3), 458–496.

Bowler, D. M. (2007). *Autism Spectrum Disorder. Psychological theory and research.* Chichester, Wiley.

Bowler, D. M., Gaigg, S. B., & Gardiner, J. M. (2008a). Subjective organisation in the free recall learning of adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders, 38*(1), 104-113.

Bowler, D. M., Gaigg, S. B., & Lind, S. (2011). Memory in autism: Binding, self and brain. In I. Roth & P. Rezaie (Eds.), *Researching the autism spectrum: Contemporary perspectives* (p. 316-346). Cambridge: Cambridge University Press.

Bowler, D.M., Gardiner, J. M., & Gaigg, S. B. (2007). Factors affecting conscious awareness in the recollective experience of adults with Asperger's syndrome. *Consciousness and Cognition, 16*(1), 124-143.

Bowler, D. M., Gardiner, J. M., & Grice, S. J. (2000a). Episodic memory and remembering in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders, 30*(4), 295-304.

Bowler, D. M., Gardiner, J. M., Grice, S. J., & Saavalainen, P. (2000b). Memory illusions: False recall and recognition in adults with Asperger's syndrome. *Journal of Abnormal Psychology, 109*(4), 663-672.

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavioural Research methods, 45*(4), 1322-1331.

Cooper, R. A., Plaisted-Grant, K. C., Baron-Cohen, S., & Simons, J. S. (2017a). Eye movements reveal a dissociation between memory encoding and retrieval in adults with autism. *Cognition, 159*. 127-138.

Cooper, R. A., Richter, F. R., Bayes, P. M., Plaisted-Grant, K. C., Baron-Cohen, S., & Simons, J. S. (2017b). Reduced hippocampal functional connectivity during episodic memory retrieval in autism. *Cerebral Cortex, 27*(2), 888-902, Crane, L., & Goddard, L. (2008). Episodic and semantic autobiographical memory in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(3). 498-506.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Frith, U. (1970a). Studies in pattern detection in normal and autistic children: I. Immediate recall of auditory sequences. *Journal of Abnormal Psychology, 76*(3), 413-420.

Frith, U. (1970b). Studies in pattern detection in normal and autistic children: II. Reproduction and production of color sequences. *Journal of Experimental Child Psychology, 10*(1), 120-135.

Fyffe, C., & Prior, M. (1978). Evidence for language recoding in autistic, retarded and normal children: A re-examination. *British Journal of Psychology, 69*(3), 393-402.

Gaigg, S. B., & Bowler, D. M. (2009). Illusory memories of emotionally charged words in autism spectrum disorder: Further evidence for atypical emotion processing outside the social domain. *Journal of Autism and Developmental Disorders*, *39*(7), 1031-1038.

Gaigg, S. B., Bowler, D. M., Ecker, C., Calvo-Merino, B., & Murphy, D.G. (2015). Episodic recollection difficulties in ASD result from atypical relational encoding: behavioural and neural evidence. *Autism Research, 8*(3), 317-327.

Gaigg, S. B., Bowler, D. M., & Gardiner, J. M. (2014). Episodic but not semantic order memory difficulties in autism spectrum disorder: Evidence from the Historical Figures Task. *Memory, 22*(6), 669-678.

Gaigg, S. B., Gardiner, J. M. & Bowler, D. M. (2008). Free recall in autism spectrum disorder: The role of relational and item-specific encoding. *Neuropsychologia, 46*(4), 983-992.

Gardiner, J. M., Bowler, D. M., & Grice, S. J. (2003). Further evidence of preserved priming and impaired recall in adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders, 33*(3), 259-269.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Appplied Psycholinguistics, 12*(3), 349-367.

Geurts, H. M. & Vissers. M. E. (2012). Elderly with autism: Executive functions and memory. *Journal of Autism and Developmental Disorders, 42*(5), 665-675..

Goldstein, G., Minshew, N. J., Allen, D. N., & Seaton, B. E. (2002). High-functioning autism and schizophrenia: A comparison of an early and late onset neurodevelopmental disorder. *Archives of Clinical Neuropsychology, 17*(5), 461-475.

Gomes, C. A., Montaldi, D., & Mayes, A. (2015). The pupil as an indicator of unconscious memory: Introducing the pupil priming effect. *Psychophysiology, 52*(6) 754-769.

Grainger, C., Williams, D. M, & Lind, S. E. (2014). Online action monitoring and memory for self-performed actions in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 44*(5), 1193-1206.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*(6), 424-429.

Haenschel, C., Bittner, R. A., Haertling, F., Rotarsk-Jagiela, A., Maurer, K., Singer, W., & Linden, D. E. J. (2007). Contribution of impaired early-Stage visual processing to working memory dysfunction in adolescents with Schizophrenia. *Archives of General Psychiatry, 64*(11), 1229-1240.

Hartmann, M., & Fischer, M. H. (2014). Pupillometry: The eyes shed fresh light on the mind. *Current Biology, 24*(7), R281-282.

Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory, 19*(4), 398-405.

Hermelin, B. & O'Connor, N. (1967). Remembering of words by psychotic and subnormal children. *British Journal of Psychology, 58*(3), 213-218.

Kamio, Y., & Toichi, M. (2007). Memory illusion in high-functioning autism and Asperger's disorder. *Journal of Autism and Developmental Disorders, 37*(5), 867-876.

Koen, J. D., & Yonelinas, A. P. (2014). The effects of healthy aging, amnestic mild cognitive impairment, and Alzheimer`s disease on recollection and familiarity: A meta-analytic review. *Neuropsychology Review, 24*(3), 332-354.

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., ..., Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports, 8*(1): 4949.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science, 7*(1), 18-27.

Lever, A. G., & Geurts, H. M. (2016). Age-related differences in cognition across the adult lifespan in autism spectrum disorder. *Autism Research, 9*(6), 666-676.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behaviour. *Journal of Autism and Developmental Disorders, 19*(2), 185-212.

Loucas, T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., & Baird, G. (2008). Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry, 49*(11), 1184-1192.

Massand, E. (2011). *Event-related potential brain correlates of episodic and semantic memory in adults with autism spectrum disorder* (Unpublished doctoral dissertation). City University London, UK.

Massand, E., & Bowler, D. M. (2015). Atypical neurophysiology underlying episodic and semantic memory in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(2), 298-315.

Massand, E., Bowler, D. M., Mottron, L., Hosein, A., & Jemel, B. (2013). ERP correlates of recognition memory in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(9), 2038-2047.

Minshew, N. J., & Goldstein, G. (1993). Is autism an amnesic disorder? Evidence from the California Verbal Learning Test. *Neuropsychology, 7*(2), 209-216.

Minshew, N. J., Goldstein, G., Muenz, L. R., & Payton, J. B. (1992). Neuropsychological functioning in nonmentally retarded autistic individuals. *Journal of Clinical and Experimental Neuropsychology*, *14*(5), 749-761.

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The “subjective” pupil old/new effect: Is the truth plain to see? *International Journal of Psychophysiology, 89*(1), 48-56.

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In J. A. Burack, T. Charman, N. Yirmiya & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131-148). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36*(1), 27-43.

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*(10), 1346-1353.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology, 83*(1), 56-64.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*(3), 560-569.

Plaisted, K, O'Riordan, M., & Baron-Cohen, S. (1998a). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry, 39*(5), 765-775.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1*(1), 125-126.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology, 60*(2), 211-229.

Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: Evidence from eye tracking. *Developmental Science, 12*(4), 549-556

Ring, M., Derwent, C. L. T., Gaigg, S. B. & Bowler, D. M. (2017). Structural learning difficulties implicate altered hippocampal functioning in adults with Autism Spectrum Disorder. *Journal of Abnormal Psychology, 126*(6), 793-804.

Ring, M., Gaigg, S. B., & Bowler, D. M. (2016). Relational memory processes in adults with Autism Spectrum Disorder. *Autism Research, 9*(1), 97-106.

Rock, P. L., Rosier, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine, 44*(10), 2029-2040.

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature reviews. Neuroscience, 10*(3), 211-223.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 156-163.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174-215.

Souchay, C., Wojcik, D. Z., Williams, H. L., Crathern, S., & Clarke, P. (2013). Recollection in adolescents with autism spectrum disorder. *Cortex, 49*(6), 1598-1609.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers, 31*(1), 137-149.

Steinhauer, S. R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. *Annals of the New York Academy of Sciences, 658*, 182-204.

Tager-Flusberg, H. (1991). Semantic processing in the free recall of autistic children: Further evidence for a cognitive deficit. *British Journal of Developmental Psychology, 9*(3), 417-430.

Tager-Flusberg, H., & Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358*(1430), 303-314.

Tanweer, T., Rathbone, C. J., & Souchay, C. (2010). Autobiographical memory, autonoetic consciousness, and identity in Asperger syndrome. *Neuropsychologia, 48*(4), 900-908.

Sapey-Triomphe, L.-A., Sonié, S., Hénaff, M.-A., Mattout, J., & Schmitz, C. (2018). Adults with autism tend to underestimate the hidden environmental structure: Evidence from a visual associative learning task. *Journal of Autism and Developmental Disorders*. [Epub ahead of print]

The Psychological Corporation (2000). *Wechsler Adult Intelligence Scale III UK edition (WAIS-III UK; 3rd ed.)*. London, UK: Author.

The Psychological Corporation (2008). *Wechsler Adult Intelligence Scale IV UK edition (WAIS-IV UK; 4th ed.)*. London, UK: Author

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1-12.

Tulving, E. (2002). Episodic memory: From mind to brain. Annual Review of Psychology, 53, 1-25.

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*(2), 167-174.

Võ, M. L.-H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology, 45*(1), 130-140.

Wilkinson, D. A., Best, C. A., Minshew, N. J., & Strauss, M. S. (2010). Memory awareness in individuals with autism. *Journal of Autism and Developmental Disorders, 40*(11), 1371-1377.

Wojcik, D. Z., Moulin, C. J. A., & Souchay, C. (2013). Metamemory in children with autism: Exploring the "feeling-of-knowing" in episodic and semantic memory. *Neuropsychology*, *27*(1), 19-27.

Table 1. *Descriptive statistics for Autism Spectrum Disorder (ASD) and typically developing (TD) individuals*

	ASD (23m, 4f)		**TD (23m, 7f)**				**Cohen's**	
Measure	***M***	***SD***	***M***	***SD***	***t*(df)**	***p***	***d***	**CI**
Age (years)	42.31	11.5	43.98	12.7	0.52 (55)	.61	0.14	-0.38, 0.66
VIQ/VCI[a]	110	15.0	113	14.0	0.72 (55)	.47	0.19	-0.33, 0.71
PIQ/PRI[b]	106	15.3	106	13.8	0.12 (55)	.91	0.03	-0.49, 0.55
FIQ[c]	110	14.8	110	13.5	0.09 (53)	.93	0.02	-0.51, 0.55
Baseline pupil	3.21	0.57	3.17	0.45	0.34 (55)	.74	0.09	-0.43, 0.61
ADOS-C[d]	2.52 (0-5)	1.4						
ADOS-RSI[e]	5.76 (1-13)	2.9						
ADOS-Total[f]	8.29 (5-17)	3.4						
ADOS-Im[g]	1.15 (0-2)	0.6						
ADOS-SB[h]	1.14 (0-5)	1.4						

Note. [a]VIQ - Verbal IQ (WAIS-III[UK]) or VCI - Verbal Comprehension Index (WAIS-IV[UK]). [b]PIQ - Performance IQ (WAIS-III[UK]) or PRI - Perceptual Reasoning Index (WAIS-IV[UK]). [c]Full-scale IQ (WAIS-III[UK] or WAIS-IV[UK]) was available for 26 ASD and 29 TD individuals. [d-h]ADOS scores are from a subset of 21 out of the 27 ASD individuals. [d]ADOS - Communication subscale. [e]ADOS - Reciprocal Social Interaction subscale. [f]ADOS Total score - Communication + Reciprocal Social Interaction. [g]ADOS - Imagination/Creativity subscale. [h]ADOS - Stereotyped Behaviours and Restricted Interests subscale. ADOS scores are presented with range in brackets.

Table 2. C*riteria for lists of pictures and words according to Snodgrass and Vanderwart (1980).*

Criterion	**List 1** *M* (*SD*)	**List 2** *M* (*SD*)	**List 3** *M* (*SD*)	**List 4** *M* (*SD*)	$F(3,40)$	p	η_p^2
Letter nr[a]	5.70 (1.16)	5.50 (1.43)	5.70 (1.34)	6.20 (1.23)	0.53	.66	.04
Name agre[b]	0.98 (0.04)	0.93 (0.09)	0.92 (0.09)	0.97 (0.04)	1.80	.17	.13
Image agre[c]	3.91 (0.45)	3.77 (0.68)	3.55 (0.50)	3.70 (0.68)	0.63	.60	.05
Familiarity	3.58 (0.79)	3.39 (0.65)	3.15 (1.07)	3.25 (0.93)	0.45	.72	.04
Complexity	2.47 (0.70)	2.73 (0.90)	3.18 (0.65)	3.01 (0.98)	1.46	.24	.11
Frequency[d]	10.30 (7.02)	10.20 (5.85)	11.50 (6.26)	11.90 (6.77)	0.17	.91	.01

Note. [a]Letter number. [b]Name agreement. [c]Image agreement. [d]Word frequency- Kucera & Francis.

Table 3. *Luminance in cd/m^2 for sets of meaningless shape images and meaningful pictures used in Experiment 1.*

	Set A *M* (*SD*)	**Set B** *M* (*SD*)	**Set C** *M* (*SD*)	**Set D** *M* (*SD*)
Shapes	117.29 (1.81)	117.42 (2.12)	-	-
Pictures	164.71 (11.87)	156.41 (8.27)	155.62 (8.54)	160.37 (15.74)

Table 4. *Means (M) and Standard deviations (SD) for Hit rates, false alarm rates and Corrected recognition rates (Hits- false alarms) for recognition (total), remember (R) and know (K) responses for words, non-words, pictures and shapes.*

	TD			ASD			Both		
	Total	**R**	**K**	**Total**	**R**	**K**	**Total**	**R**	**K**
	M (***SD***)	***M*** (***SD***)	***M*** (***SD***)	***M*** (***SD***)	***M*** (***SD***)	***M*** (***SD***)	***M*** (***SD***)	***M*** (***SD***)	***M*** (***SD***)
Hits									
Words	0.74 (0.22)	0.56 (0.29)	0.18 (0.20)	0.58 (0.28)	0.39 (0.26)	0.19 (0.16)	0.66 (0.26)	0.48 (0.28)	0.19 (0.18)
Non-words	0.77 (0.24)	0.66 (0.25)	0.11 (0.14)	0.64 (0.27)	0.45 (0.26)	0.20 (0.19)	0.71 (0.26)	0.56 (0.27)	0.15 (0.17)
Pictures	0.90 (0.16)	0.75 (0.25)	0.16 (0.20)	0.83 (0.16)	0.59 (0.26)	0.24 (0.21)	0.87 (0.16)	0.67 (0.26)	0.19 (0.20)
Shapes	0.70 (0.24)	0.62 (0.24)	0.08 (0.09)	0.58 (0.27)	0.46 (0.31)	0.12 (0.13)	0.64 (0.26)	0.54 (0.29)	0.10 (0.11)
False alarms									
Words	0.08 (0.10)	0.04 (0.09)	0.04 (0.06)	0.13 (0.16)	0.06 (0.10)	0.07 (0.11)	0.10 (0.14)	0.05 (0.09)	0.05 (0.08)
Non-words	0.07 (0.12)	0.03 (0.08)	0.03 (0.07)	0.14 (0.17)	0.06 (0.10)	0.08 (0.11)	0.10 (0.15)	0.05 (0.09)	0.05 (0.09)
Pictures	0.06 (0.11)	0.03 (0.07)	0.03 (0.04)	0.06 (0.10)	0.02 (0.04)	0.04 (0.07)	0.06 (0.10)	0.03 (0.06)	0.03 (0.06)
Shapes	0.11 (0.15)	0.07 (0.11)	0.04 (0.09)	0.17 (0.19)	0.09 (0.13)	0.08 (0.12)	0.14 (0.17)	0.08 (0.11)	0.06 (0.11)

	TD			ASD			Both		
	Total	**R**	**K**	**Total**	**R**	**K**	**Total**	**R**	**K**
	M	***M***	***M***	***M***	***M***	***M***	***M***	***M***	***M***
	(*SD*)	**(*SD*)**	**(*SD*)**	**(*SD*)**	**(*SD*)**	**(*SD*)**	**(*SD*)**	**(*SD*)**	**(*SD*)**
Corrected recognition (Hits –False Alarms)									
Words	0.66	0.52	0.14	0.45	0.33	0.12	0.56	0.43	0.13
	(0.27)	(0.30)	(0.19)	(0.27)	(0.24)	(0.20)	(0.29)	(0.29)	(0.19)
Non-words	0.71	0.63	0.08	0.51	0.39	0.12	0.61	0.51	0.10
	(0.26)	(0.27)	(0.13)	(0.25)	(0.23)	(0.21)	(0.27)	(0.28)	(0.17)
Pictures	0.85	0.72	0.13	0.77	0.57	0.20	0.81	0.65	0.16
	(0.23)	(0.27)	(0.20)	(0.19)	(0.26)	(0.22)	(0.22)	(0.27)	(0.21)
Shapes	0.59	0.55	0.04	0.41	0.37	0.04	0.51	0.46	0.04
	(0.27)	(0.25)	(0.12)	(0.33)	(0.33)	(0.14)	(0.31)	(0.30)	(0.13)

Table 5

Means (M) and Standard Deviations (SD) for A' (sensitivity) and B" (response bias) for Remember responses for words, non-words, pictures, and shapes for ASD and TD groups.

	ASD	**TD**	**Total**
	***M* (*SD*)**	***M* (*SD*)**	***M* (*SD*)**
A'			
Words	0.78 (0.13)	0.86 (0.10)	0.82 (0.12)
Non-words	0.79 (0.14)	0.89 (0.12)	0.84 (0.14)
Pictures	0.88 (0.10)	0.92 (0.10)	0.90 (0.10)
Shapes	0.76 (0.20)	0.86 (0.14)	0.81 (0.18)
B"			
Words	0.47 (0.36)	0.58 (0.26)	0.52 (0.32)
Non-words	0.48 (0.34)	0.54 (0.38)	0.52 (0.36)
Pictures	0.60 (0.27)	0.45 (0.40)	0.52 (0.35)
Shapes	0.39 (0.38)	0.50 (0.33)	0.45 (0.36)

Figures

Figure 1. Examples of materials, from top to bottom- words, non-words, pictures, shapes.

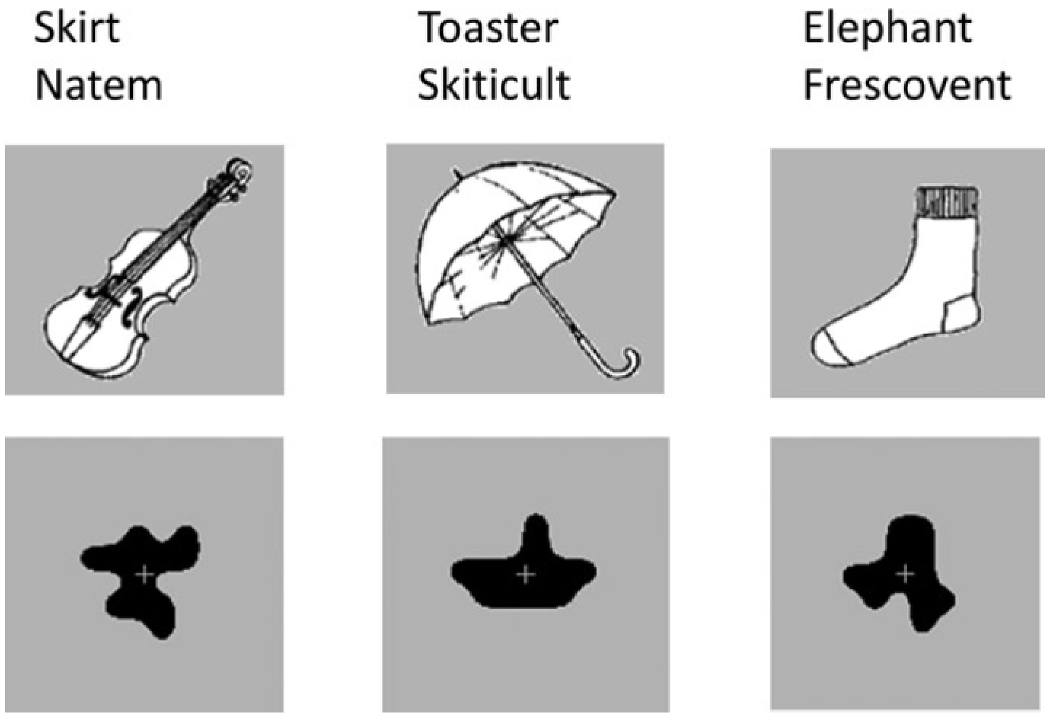


Figure 2. Procedures for study (top) and test phase (bottom).

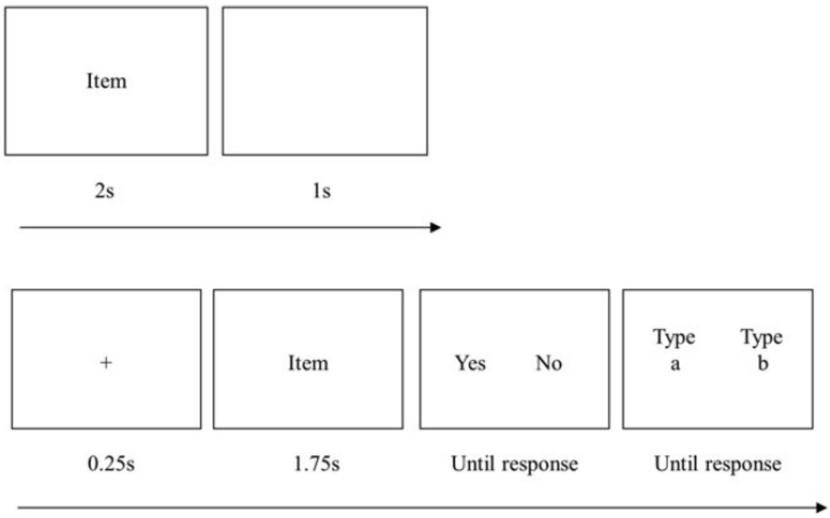


Figure 3. Corrected recognition (Hits minus false alarms) for Remember and Know collapsed across modalities and meaningfulness for the two groups (left). Corrected recognition (Hits minus false alarms) for verbal and visual materials split up by meaning collapsed across the two groups (right). Error bars are standard error of the mean.

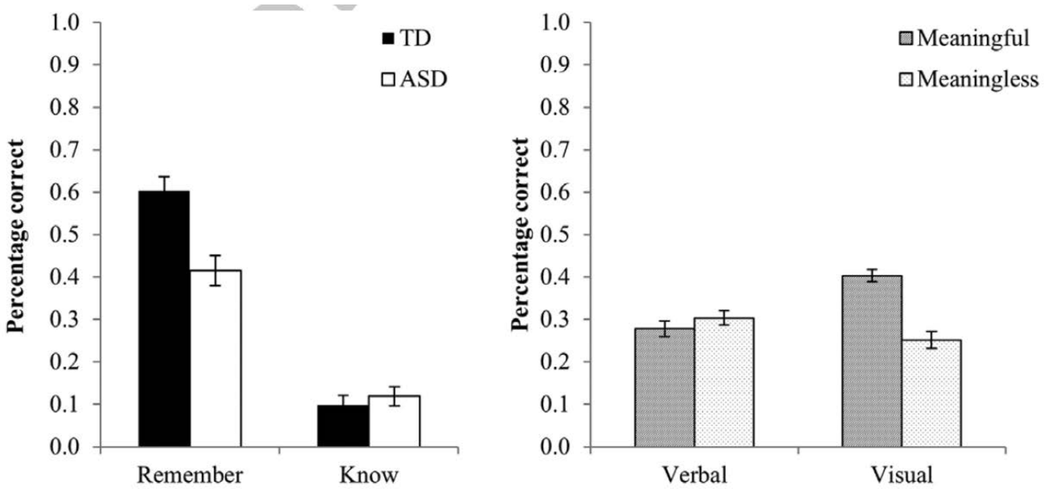


Figure 4. Maximum pupil dilation ratio (pupil size during task/ pupil size during baseline) for old and new items collapsed across material and meaning for the two groups during retrieval. Error bars are standard error of the mean.

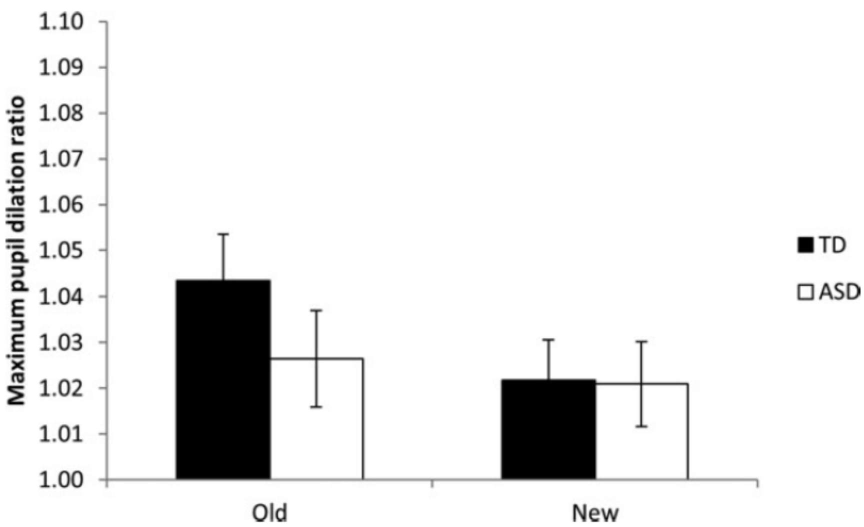


Figure 5. Association between corrected recognition rates and the maximum pupil dilation ratio (pupil size during task/pupil size during baseline) between old and new items. The correlation illustrates that larger pupils in response to old vs. new items was related to higher corrected recognition rates in the behavioural response.

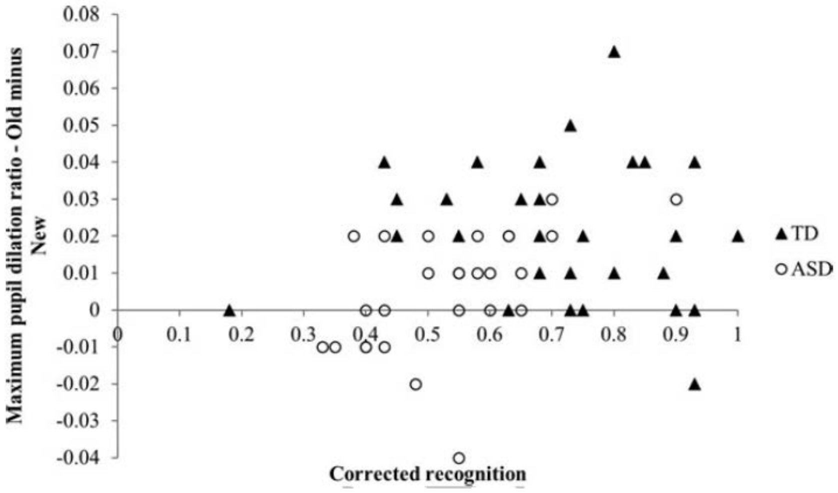
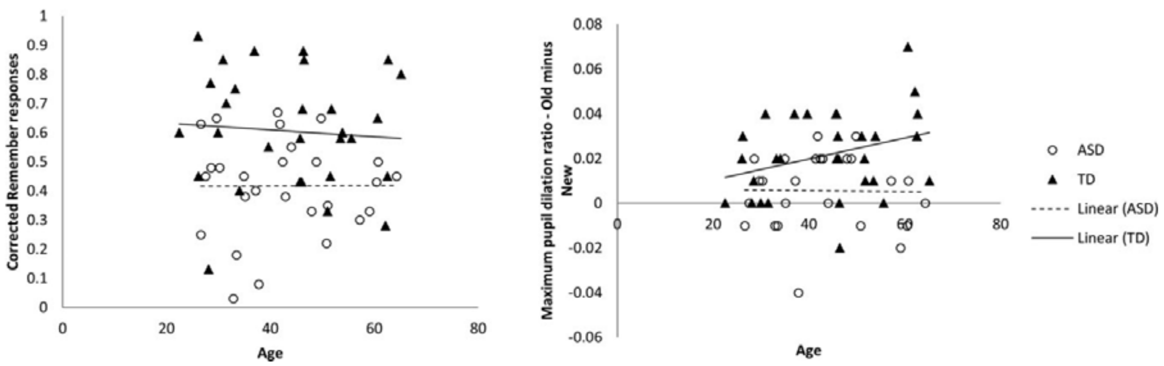


Figure 6. Association between corrected Remember rates and age (left) and maximum pupil dilation ratio (pupil size during task/pupil size during baseline) between old and new items and age (right). Age effects were larger for the TD as opposed to the ASD group on corrected Remember responses and pupil Old-New ratio.



Acknowledgements

The presentation at the International Meeting for Autism Research 2015 in Salt Lake City, Utah, USA in 2015 was supported by a student travel award from the International Society for Autism Research (INSAR) to the first author. The first author was supported by a PhD studentship from City, University of London. We are very grateful to Dr Corinna Haenschel for providing the shape stimuli used in this experiment. We would like to thank all participants for taking part in this research. There are no conflicts of interest for any of the authors.