

# Referential gaze and word learning in adults with autism

Autism  
2015, Vol. 19(8) 944–955  
© The Author(s) 2014  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/1362361314556784  
aut.sagepub.com



Iyad Aldaqre, Markus Paulus and Beate Sodian

## Abstract

While typically developing children can use referential gaze to guide their word learning, those with autism spectrum disorder are often described to have problems with that. However, some researchers assume that the ability to follow gaze to select the correct referent can develop in autism later compared to typically developing individuals. To test this assumption, we compared the performance of adults with and without autism on a word learning task while recording their gaze behavior using an eye tracker. Results showed that both groups mostly chose the correct referent, but less so for the autism spectrum disorder group when the distractor's saliency was increased, suggesting that the ability to learn novel words by referring to gaze develops in autism spectrum disorder, but not fully, relative to their typically developing peers.

## Keywords

autism spectrum disorder, development, eye tracking, referential gaze, word learning

## Introduction

When people hear a speaker uttering a new word, one very important strategy they use to determine the intended referent is relying on the speaker's direction of gaze (e.g. Baldwin, 1991). By the age of 2 years, infants can already use gaze information actively to learn novel word–object associations (Baldwin, 1993; Houston-Price et al., 2006; Moore et al., 1999; Paulus and Fikkert, 2014). The skill of following someone's gaze to attend to the same location, also known as joint attention, develops even earlier in life, around the end of the first year (Paulus, 2011; Tomasello, 2006). However, when joint attention skills are disrupted, like in the case of autism spectrum disorder (ASD), word mapping errors arise (Akechi et al., 2011; Baron-Cohen et al., 1997; Preissler and Carey, 2005). For example, Baron-Cohen et al. (1997) tested children with autism with a word learning paradigm, in which the experimenter presented two novel objects to the children and attempted to teach them the name of one object (the target) by looking at it while uttering its name. Children with autism failed to move their attention to the same location as the experimenter's, and they attributed the novel name to the object they were attending to at the time, a strategy known as the Listener's Direction of Gaze (LDG). Also, in situations where a salient distractor is presented simultaneously with the target, children at risk of autism were not able to learn

the new word–object association correctly (Gliga et al., 2012), while typically developing children can do that by the end of the second year of life (Moore et al., 1999).

Although many studies suggest impaired ability to follow gaze in autism, as mentioned above, others have shown that people with ASD can in fact follow gaze correctly (Chawarska et al., 2003; Kuhn et al., 2010; Senju et al., 2004). This suggests that the problem in ASD with learning novel words in a social context cannot be explained by the mere inability to follow gaze. Rather, their decreased preference of the target object can be explained by the inability to perceive the object being looked at by the speaker as special, and therefore not appreciating that it is relevant to what the speaker is saying (Akechi et al., 2011; Baron-Cohen et al., 1985, 1997; Gliga et al., 2012; Waxman and Gelman, 2009; but see Gillespie-Lynch et al., 2013). This account is supported by the previous literature, which demonstrated that children with autism, unlike typically

---

Ludwig-Maximilians-University, Germany

### Corresponding author:

Iyad Aldaqre, Department of Psychology, Ludwig-Maximilians-University, Leopoldstr. 13, 80802 Munich, Germany.  
Email: iyad.aldaqre@campus.lmu.de

developing children, process gaze cues in a similar way to non-social cues, like non-biological eyes or arrows (Chawarska et al., 2003; Greene et al., 2011; Senju et al., 2004). Moreover, adults with ASD do not seem to interpret gaze cues as indicators for relevant information (Böckler et al., 2014). Consequently, one could argue that people with autism have problems in understanding the referential nature of human eye gaze.

It is important to note that not all children with autism show atypical performance, and they are able to use gaze direction to learn the correct word–object association (Akechi et al., 2011; Luyster and Lord, 2009). Gliga et al. (2012) have also shown that children at risk of autism with preserved social and communicative skills can rely on the direction of gaze of an actor to learn new words, even when a distractor is more salient than the target object. From these findings, the authors suggested that children with ASD may develop the ability to use the speaker’s direction of gaze to learn a novel word–object association; however, it might be delayed, relative to children without autism (Akechi et al., 2011; Gliga et al., 2012; Luyster and Lord, 2009). If this is true, it would not be the only ability that is delayed in ASD. While they are often described to fail the theory of mind (ToM) tasks (e.g. Baron-Cohen et al., 1985), children with autism, who have a higher verbal mental age, were shown to pass these tasks (Happé, 1995). These findings lend some support to the assumption that social-cognitive development in people with ASD is delayed and that adults with autism might be able to understand the referential nature of another’s gaze cue. Consequently, it would be important to assess whether the inability to use another’s gaze cue in a word learning situation—as reported from children at risk of autism (Gliga et al., 2012)—constitutes an enduring problem or whether persons with autism become able to do so later in development. Findings that adults with ASD interpret gaze cues differently than typically developing persons (Böckler et al., 2014) provide preliminary evidence for the first claim, whereas findings that some children with autism develop the ability to use gaze cues (Akechi et al., 2011) provide preliminary support for the latter claim.

Moreover, it should be noted that recent findings demonstrated a differentiation between explicit and implicit forms of social-cognitive abilities (e.g. Frith and Frith, 2008, 2012). For example, in the ToM research tradition, researchers noted that participants with autism, who have higher verbal abilities, are able to demonstrate ToM competencies in explicit tasks—that is, when they are verbally asked to explicitly reason about another’s belief (Happé, 1995). In contrast, implicit measures of their ToM understanding, often assessing their looking behavior in eye-tracking paradigms, indicate persisting deficits in their ToM competencies (Senju et al., 2009, 2010). However, other studies have demonstrated reversed results with respect to other social-cognitive competencies. For example, a recent study

demonstrated intact implicit, but impaired explicit level 1 perspective-taking in adults with autism (Schwarzkopf et al., 2014). Given this intermixed picture of results, it would be interesting to assess whether or not implicit and explicit measures converge in the assessment of word learning abilities in people with autism.

To examine these issues, we employed eye-tracking technology and used a computerized version of a word learning task to assess word learning from gaze cues in adults with autism. In this task, participants were presented with unfamiliar objects and an animated face that looked at one of the objects while teaching the participants a novel word. Subsequently, participants were administered two types of test trials: explicit trials in which they were asked to select the target object from a set of cards and implicit test trials in which we employed a preferential looking paradigm. This allowed us to assess whether or not there is any dissociation between implicit and explicit responses. Given the findings that people with autism might use gaze cues, but process them—unlike typically developing people—in the same manner as nonsocial cues (e.g. arrows), we introduced a second condition. In this condition, one of the objects was cued by gaze during the labeling action, while the other object was provided with a (nonsocial) saliency cue (see Moore et al., 1999). This situation examined whether participants rather rely on the social or the nonsocial cue in their word learning, as both cues were presented in conflict at the same time. Therefore, it is a stricter test of the ability to use direction of gaze to determine an intended referent of a novel word and a more thorough assessment of their understanding of the referential nature of another’s gaze.

## Methods

### Participants

The final sample included 15 high-functioning adults with ASD aged 19–61 years (6 females; mean age: 36.9 years) and 15 neuro-typical (NT) adults aged 20–53 years (9 females; mean age: 32.5 years). Adults with ASD were diagnosed by a qualified clinical psychologist or psychiatrist, and they met the International Classification of Diseases 10th Revision (ICD-10) criteria for Asperger syndrome ( $N=8$ ), autistic disorder ( $N=4$ ), or childhood autism ( $N=3$ ). Four additional participants were excluded from the analyses due to refusal to continue the session (1 ASD), technical problems with the experimental procedure (1 ASD and 1 NT), or later change in the diagnosis (1 ASD). All participants completed the German shortened version of the autism quotient (AQ-k, Freitag et al., 2007; originally developed by Baron-Cohen et al., 2001). Other measures included the Culture Fair Test 20-R (CFT 20-R) for non-verbal intelligence (Weiss, 2006) and the German vocabulary test (Mehrfachwahl-Wortschatz-Intelligenztest

**Table 1.** Means, SD, and range of the age; autism quotient (AQ-k); non-verbal intelligence (CFT 20-R); and verbal intelligence (MWT-B) of participants.

	ASD ( <i>n</i> = 15)		NT ( <i>n</i> = 15)		<i>p</i> value
	<i>M</i> (SD)	Range	<i>M</i> (SD)	Range	
Age (years)	36.9 (11.6)	19–61	32.5 (11.1)	20–53	ns
AQ-k	24.5 (5.9)	9–31	5.5 (4)	1–14	<i>p</i> < 0.001
CFT 20-R	107.1 (26.6)	65–149	108.3 (14.3)	92–130	ns
MWT-B	119.1 (13.3)	100–136	121.5 (13.4)	95–143	ns

AQ-k: autism quotient–short version; CFT 20-R: Culture Fair Test 20-R; MWT-B: Mehrfachwahl-Wortschatz-Intelligenztest; ASD: autism spectrum disorder; NT: neuro-typical; SD: standard deviation; ns: not significant. Significant differences were observed only in the AQ-k score ( $t(28) = 10.3, p < 0.001$ ).

(MWT-B); Lehl, 2005) for verbal intelligence. Demographic data of the participants are presented in Table 1. Participants gave a written consent before starting the experiment and were given monetary compensation for their participation. All participants had normal or corrected-to-normal vision. The mother tongue of all participants was German, except for one control participant, who spoke German fluently.

### Stimuli

Stimuli were short animation movies in which a cartoon actress taught the participants a novel word–object association. Three conditions were presented to each participant: one familiarization and two test conditions, each of which was presented twice. In the familiarization condition, four different well-known objects (an apple, a car, a fish, and a boat) were presented on the computer display, and participants were asked to look at a specific one (e.g. the apple). The test conditions were similar to the “static control” and the “mismatch” conditions described by Moore et al. (1999). Each test condition was divided into two trials: learning and response trials (see procedure for a detailed description), resulting in 10 trials for each participant in total. In the learning trial of the static condition, the actress was presented with two novel objects in front of her, and she looked at and labeled one of them with a novel name. Figure 1(a) shows an example of the learning trial, with the areas of interest (AOIs) from which the gaze data were exported. The learning trial of the mismatch condition was similar to the static condition, except that the distractor object started jiggling when the actress looked at and labeled the target object. The response trials of both test conditions were similar to the familiarization trial, except that the objects were the two previously presented objects during the learning trial of each condition and two additional distractors.

Following each of the three conditions, a set of four cards, with the previously presented objects printed on them, was handed to the participants, and they were asked to explicitly select the previously labeled object (i.e. the

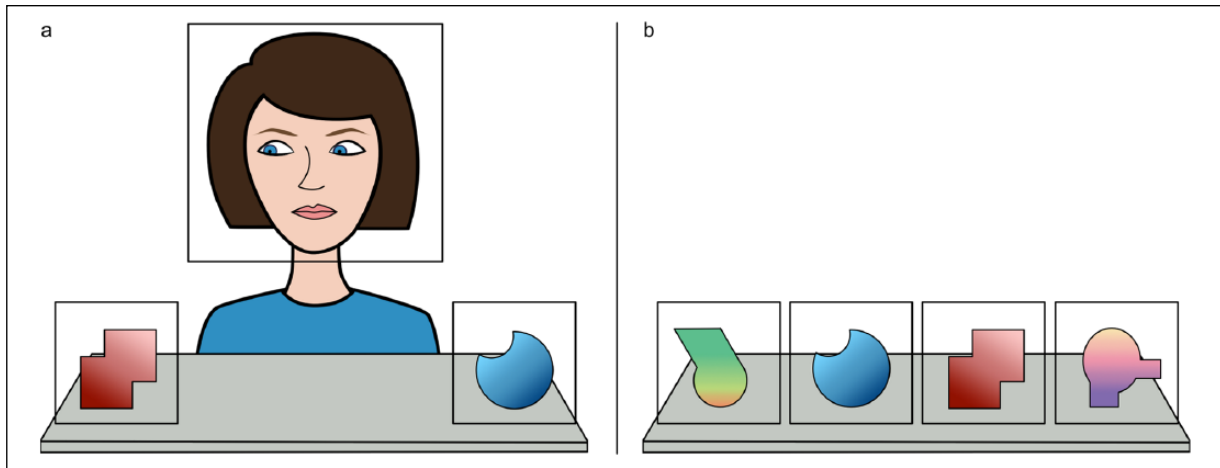
target) to assess whether the new word was learned and could be used in an interactive situation. No feedback was given to the participant about their choice, nor about their looking behavior.

All objects used in the movies were digitally scanned from the German language development test for 3- to 5-year-olds (Sprachentwicklungstest für drei-bis fünfjährige Kinder - SETK 3–5; Grimm, 2001). The pictures of known objects, which were used in the familiarization trial, were an apple, a car, a fish, and a boat. The novel objects, which were used in the test trials with their corresponding names, were from the fantasy-words subtest of the SETK 3–5, and their names were standardized for the German language.

### Apparatus and procedure

Participants sat on a height-adjustable office chair, approximately 60 cm away from the eye tracker. Gaze data were recorded with a Tobii T60 eye tracker (Tobii Technology, Sweden) at 60 Hz sampling rate. The stimuli were presented on the 17-in display integrated into the eye tracker. Both stimulus presentation and data acquisition were done using the Tobii Studio software (Tobii Technology).

At the beginning of the session, participants were seated in front of a table and were asked to sign the written consent and to fill in some demographic information. Then, they were instructed to simply sit in front of the eye tracker and watch some animated movies. No further instructions were given. At the beginning of the videos, an animated, two-dimensional cartoon actress was presented on the display, with a small tabletop in front of her; she greeted the participant and introduced herself. Afterward, the actress disappeared and the familiarization started. In the familiarization, four familiar objects were presented, and after a 4-s period, the actress asked the participant to look at one of the objects. For example, she would say in German “Look! The apple!” The 4-s period at the beginning of the trial was included as a baseline to control for saliency and novelty effects on the looking duration at the objects. Four seconds after the sentence had finished, the objects disappeared and the condition was repeated again with shuffled



**Figure 1.** Example pictures of the stimuli of (a) the learning trial and (b) the response trial with the AOIs overlaid on the face and objects. Different objects were produced for illustration purposes; the original objects from the SETK 3–5 are not presented due to copyright issues.

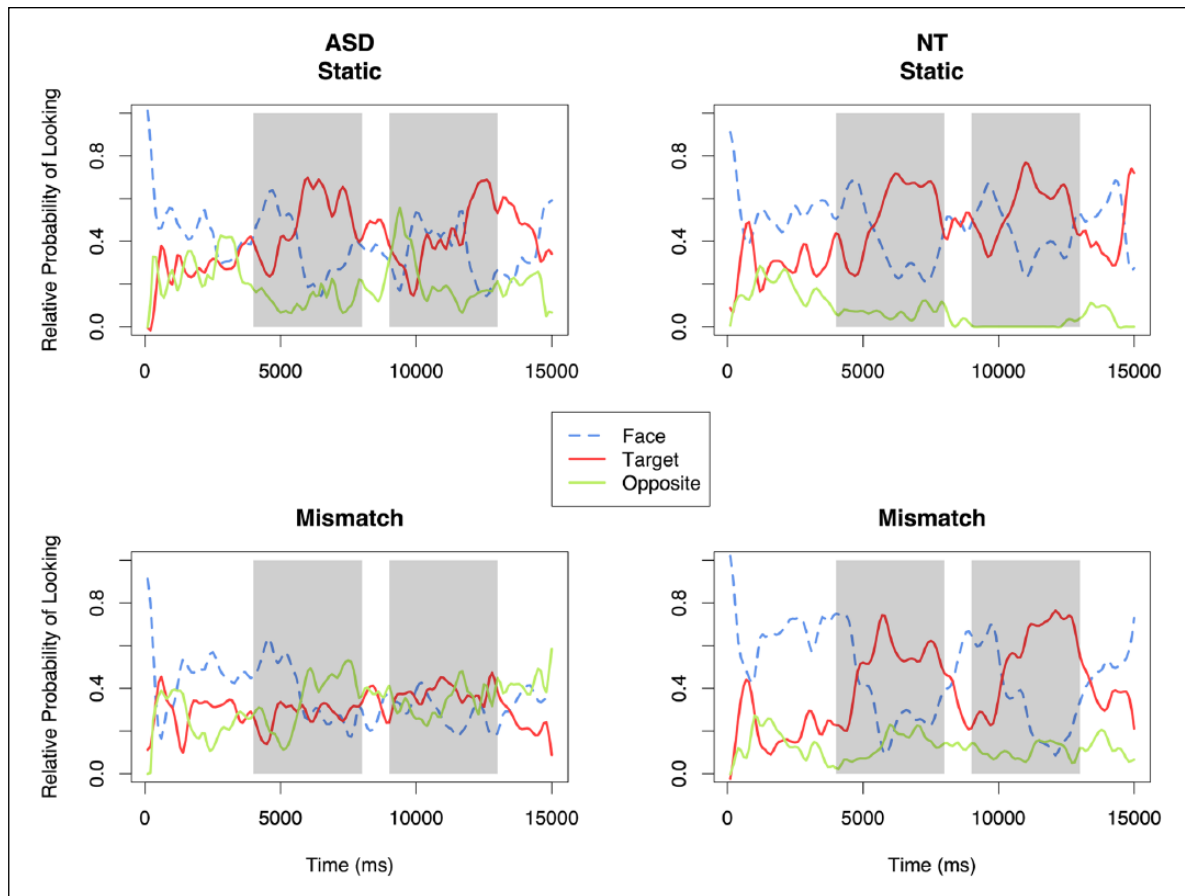
object locations. Following the repetition of the familiarization condition, a black screen was presented and the participant was handed a set of cards, with the previously viewed objects printed on them, and was asked to select the target object and give it to the experimenter. After the explicit response was finished, the static test condition started. In the learning trial of the static condition, the actress was presented, looking straight at the participant, with two novel objects in front of her. Approximately 3 s from trial onset, she looked at one of the objects and called it with a novel name and then looked back at the participant. For example, she would say in German “That is a plarte.” The actress repeated the labeling action two times, each of which lasted approximately 5 s. Following the learning trial, the test trial of the static condition was presented (see Figure 1(b) for example), which was similar to the familiarization condition in procedure. The two objects from the learning trial were presented (i.e. the target and the opposite objects), in addition to two novel distractors, and the actress was not visible. Four seconds from trial onset, the actress asked the participant to look at the target. Four seconds after the sentence had finished, the objects disappeared and the whole test condition was repeated with shuffled object locations. After the static condition was repeated, a black screen was presented and the experimenter gave the cards to the participant to select the target object. Then, the mismatch condition started. The mismatch condition was identical to the static condition, consisting of one learning and one test trials, with two main differences: different novel objects were used, and the distractor in the learning trial (i.e. the opposite object) started jiggling while the actress looked at the target and labeled it. This increase in saliency of the opposite object was employed as a second, conflicting cue in the trial with the gaze cue of the actress. After the mismatch condition

was repeated, a black screen was presented and the experimenter gave the cards to the participant to select the target object. Following the explicit response of the participant, the experiment ended. The assignment of the target object in each condition was counterbalanced between participants to control for the physical characteristics of the objects. The order of the conditions remained fixed between participants to avoid affecting their spontaneous looking pattern. If the mismatch condition was presented before the static condition, participants might have looked longer to the distractor in the static condition, expecting it to move. After the experiment was finished, participants were asked to sit again in front of the table to do the AQ test and the verbal and non-verbal intelligence tests. In some cases, the control measures were administered before the experiment starts, while the experimental equipment was prepared.

### Data analyses

Fixations were identified using a velocity-based filter (Salvucci and Goldberg, 2000). A fixation was defined as all consecutive gaze samples with a velocity of about 52 deg/s or less and at least 80 ms in duration. All data pre-processing and analyses were done using the statistical computing language “R” (R Core Team, 2013) and some of its packages (“aspace”: Bui et al., 2012; “ez”: Lawrence, 2013; “reshape2”: Wickham, 2007; “zoo”: Zeileis and Grothendieck, 2005).

**Learning trials.** Data were analyzed from three AOIs, one for each of the two objects and one for the actress’ face (see Figure 1(a) for example). To assess whether the ASD group looked less to the face of the actress compared to the NT group, absolute looking time to the face of the



**Figure 2.** Probability of looking at each of the objects and the face during learning trials for the ASD and the NT groups in both conditions. The shaded areas represent the periods during which the actress looked at and labeled the target object. ASD: autism spectrum disorder; NT: neuro-typical.

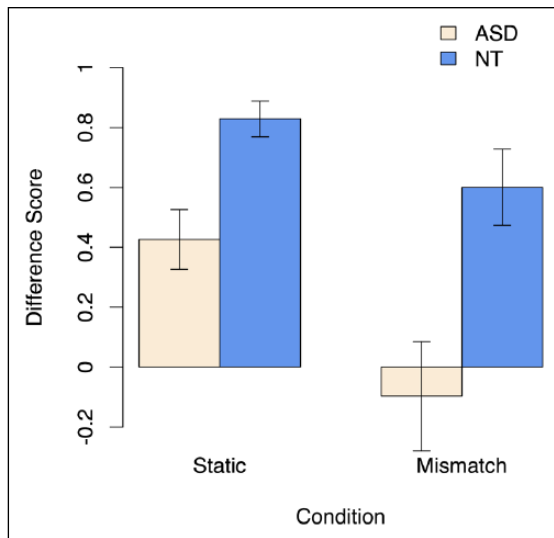
actress during the whole trial was compared between the two groups by means of a two-way analysis of variance (ANOVA), with the within-subject factor Condition (static and mismatch; see the “Apparatus and Procedure” section) and the between-subject factor Group (ASD and NT).

A difference score (DS) was calculated for looking time on the other two AOIs (i.e. the novel objects) during both labeling segments of each learning trial. This was done by subtracting the looking time to the distractor from the looking time to the target and dividing the result by the total looking time to both objects (cf. Akechi et al., 2011). The resulting value ranges from 1 (looking only at the target) to  $-1$  (looking only at the distractor). The DS was used to assess whether participants looked more to the target object when it was looked at by the actress during the learning trials. It was entered as the dependent variable in a two-way ANOVA, with the within-subject factor condition and the between-subject factor group. The between-subject factor “Gender” showed no main effect on DS nor did it interact with the other factors in the initial analysis (all  $F_s \leq 2.6$ , all  $p_s \geq 0.16$ ) and therefore was removed

from the analysis. Further analyses were carried to examine whether looking time to face correlated with DS and whether the DS differed from zero by means of a one-sample  $t$ -tests.

To have a clearer look at each group’s looking pattern over time during learning trials, the relative probability of looking at each of the three AOIs (i.e. face, target, and opposite) was calculated (cf. Bergmann et al., 2012). This was done by splitting each trial into 100 ms time bins and dividing the number of fixations to each AOI by the total number of fixations to all AOIs (see Figure 2). The average relative probability of looking at each AOI during the time in which the actress looked at and labeled the target object was calculated for each participant and each condition separately. This measure was then analyzed by means of an ANOVA, with the within-subject factors Condition and AOI, and the between-subject factor Group.

**Familiarization and test trials.** Gaze data were analyzed from the 4-s segments at the beginning of the trials (baseline) and the 4-s segments after the name of the target has ended (response segment). Four AOIs were assigned, one



**Figure 3.** Means of the difference scores (DSs) during the labeling trials for the ASD and the NT groups in both test conditions. Error bars indicate the standard error of the mean (SEM).

ASD: autism spectrum disorder; NT: neuro-typical.

for each of the objects (e.g. see Figure 1b). Trials in which there were no gaze data at any of the AOIs were omitted from analyses. Relative looking time to the objects was used as an implicit measure of word learning. It was calculated by dividing looking time on each of the AOIs by the total looking time on all AOIs and then averaged across the two repetitions of each condition for every participant. Following previous studies examining word learning and object processing (Houston-Price et al., 2006; Paulus and Fikkert, 2014; Wu and Kirkham, 2010), relative looking was used for the analyses because we were interested in the relative preference of objects, rather than the absolute looking time at the objects. A three-way ANOVA was used to analyze relative looking to the objects in the familiarization and response trials, with the within-subject factors Condition (familiarization, static, and mismatch) and AOI (target, opposite, Distractor 1, and Distractor 2) and the between-subject factor Group (ASD and NT). For this analysis, all  $p$ -values were corrected using Greenhouse–Geisser epsilon due to the violation of the sphericity assumption. The between-subject factor “Gender” showed no main effect on relative looking time nor did it interact with the other factors in the initial analysis (all  $F_s \leq 1.6$ , all  $p_s \geq 0.14$ ) and therefore was removed from the analysis.

Relative looking time to the AOIs during the baseline was subtracted from the looking time during the response segment to create a baseline DS. This score was used to indicate whether participants looked more at the target object after its name was spoken and did not prefer it for its physical properties or other characteristics. The baseline DS was then analyzed using a three-way ANOVA, with the

within-subject factors Condition and AOI and the between-subject factor Group. For this analysis, all  $p$ -values were corrected using Greenhouse–Geisser epsilon due to the violation of the sphericity assumption. Then, one-sample  $t$ -tests were used to assess whether the baseline DS significantly differed from zero for each AOI in each condition and for each group.

The correlation between looking time to face in the learning trial of each condition and relative looking time to the target in the response segments of the same condition was assessed. Additionally, correlations between DS in the learning trial of each condition and relative looking time to the target in the response segments of the same condition were assessed.

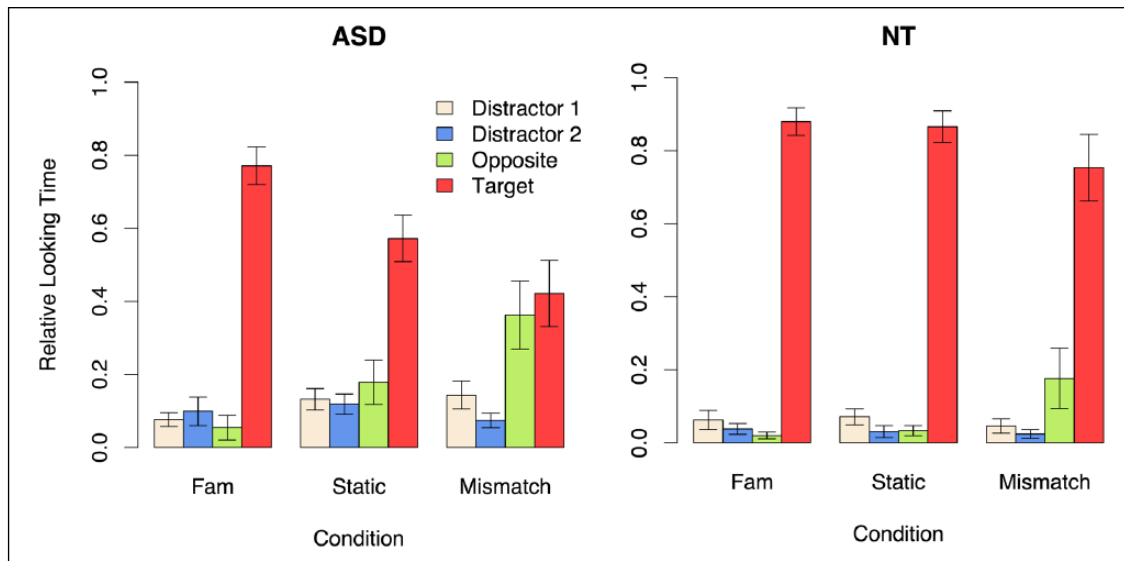
**Explicit responses.** The number of participants who selected the correct card after each condition was compared between groups for each condition by means of a chi-square test. To examine whether the proportion of participants who selected the correct card differed from chance, exact binomial tests were carried out for each condition and each group. Chance level was set to 25%, as there were four possible objects to choose from. However, because only one of the three additional items was presented in the learning trials as a possible distractor, the exact binomial tests were repeated with chance level set to 50%. The correlation between relative looking time to the target object, as an index to implicit performance, and the explicit response was assessed by means of a point-biserial correlation. Additionally, we have tested the correlation between looking time to face in the learning trial of each condition and the explicit response in that condition.

## Results

### Learning trials

The analyses of looking time to the face of the actress during the whole learning trials showed no significant main effects nor interactions (all  $F_s \leq 2.15$ , all  $p_s > 0.1$ ). When the DS for each group on both test conditions (Figure 3) was analyzed by means of an ANOVA, a significant main effect of group was found ( $F(1, 28) = 14.13$ ,  $p < 0.001$ ,  $\eta^2 = 0.26$ ), showing that the ASD group had overall lower DS than the NT group ( $t(58) = -4.1$ ,  $p < 0.001$ , Cohen’s  $d = -1.07$ ). Additionally, there was a significant main effect of condition ( $F(1, 28) = 13.8$ ,  $p < 0.001$ ,  $\eta^2 = 0.14$ ), showing that participants had higher DS in the static than in the mismatch condition ( $t(29) = 3.65$ ,  $p < 0.005$ , Cohen’s  $d = 0.67$ ). The interaction effect between the two factors (group and condition) did not reach significance ( $F(1, 28) = 2.14$ ,  $p > 0.1$ ). One-sample  $t$ -tests showed that DS is significantly different from zero in all conditions (all  $p_s < 0.001$ ), except for the mismatch condition in the ASD group ( $t(14) = 0.53$ ,  $p = 0.6$ ). No significant correlations





**Figure 4.** Means of relative looking time on the four AOIs during the response segments for the ASD and the NT groups. Error bars indicate the standard error of the mean (SEM).

AOIs: areas of interest; ASD: autism spectrum disorder; NT: neuro-typical.

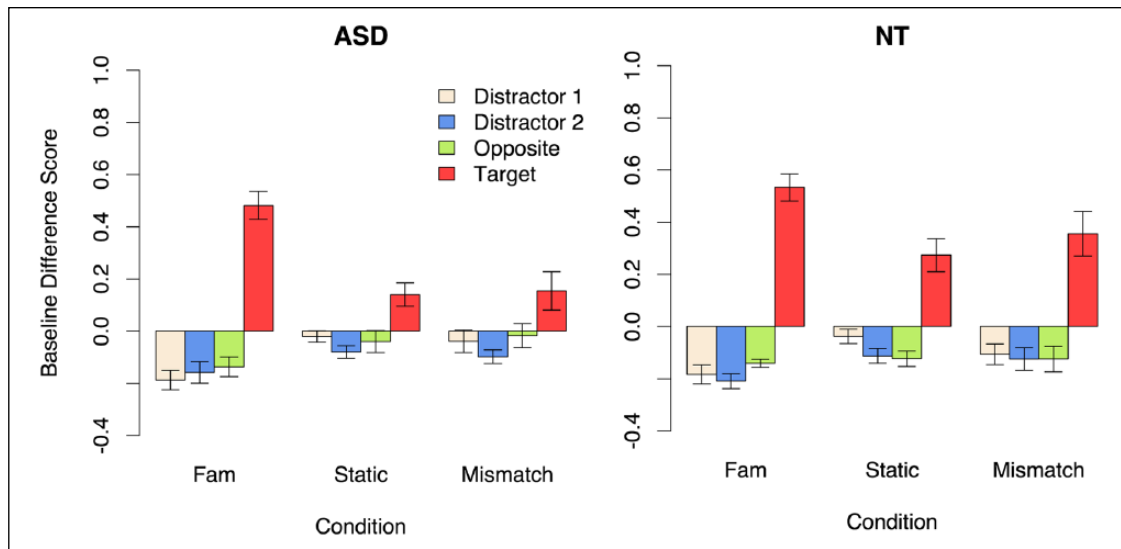
were observed between looking time to face and DS (all  $r_s < \pm 0.33$ ,  $p_s > 0.2$ ).

The analyses of relative probability of looking revealed a significant main effect of AOI ( $F(2, 56) = 11.76$ ,  $p < 0.001$ ,  $\eta^2 = 0.25$ ), showing that participants were overall more probable to look at the target object than the opposite object ( $t(59) = 5.23$ ,  $p < 0.001$ , Cohen's  $d = 1.21$ ). A significant interaction between group and AOI was also found ( $F(2, 56) = 3.97$ ,  $p < 0.05$ ,  $\eta^2 = 0.1$ ). To explore this interaction, independent samples  $t$ -tests were used to compare the probability of looking on each AOI between the two groups. These comparisons revealed that the NT group was more probable to look at the target object than the ASD group ( $t(58) = 2.25$ ,  $p < 0.05$ , Cohen's  $d = 0.58$ ), while the ASD group was more probable to look at the opposite object ( $t(58) = 3.54$ ,  $p < 0.001$ , Cohen's  $d = 0.92$ ). Additionally, there was a significant interaction between condition and AOI ( $F(2, 56) = 5.92$ ,  $p < 0.005$ ). Further analysis of this interaction revealed that participants were more likely to look at the opposite object in the mismatch than in the static condition ( $t(58) = 2.2$ ,  $p < 0.05$ , Cohen's  $d = 0.56$ ). Main effects of the remaining factors and other interactions did not reach significance (all  $F_s \leq 1.4$ , all  $p_s > 0.2$ ).

### Familiarization and test trials

Figure 4 shows the means of relative looking time on the four AOIs in all conditions for both groups during the response segment. The ANOVA showed a significant main effect of AOI on relative looking time ( $F(3, 84) = 123.63$ ,  $p < 0.001$ ,  $\eta^2 = 0.69$ ), showing that participants looked

overall more at the target compared to all other AOIs (all  $p_s < 0.001$ ). A significant interaction between AOI and group was found ( $F(3, 84) = 8.9$ ,  $p < 0.005$ ,  $\eta^2 = 0.14$ ). Paired samples  $t$ -tests were used to explore this interaction in greater detail by comparing relative looking time to each AOI between groups. All comparisons yielded a significant difference between the groups (all  $p_s < 0.05$ ), showing that the NT group looked significantly longer to the target than the ASD group, while the ASD group looked longer to the other AOIs than the NT group. A significant interaction between condition and AOI was also found ( $F(6, 168) = 8.58$ ,  $p < 0.001$ ,  $\eta^2 = 0.13$ ). Paired samples  $t$ -tests were used to explore this interaction in greater detail by comparing relative looking time to each AOI between the conditions. There was a significant difference in relative looking time to the target between the familiarization and the mismatch condition ( $t(58) = 3.1$ ,  $p < 0.005$ , Cohen's  $d = 0.79$ ), showing that participants looked significantly longer to the target object in the familiarization condition. Additionally, there was a significant difference in relative looking time to the opposite object between the familiarization and mismatch conditions ( $t(58) = 3.5$ ,  $p < 0.001$ , Cohen's  $d = 0.91$ ) and between the static and the mismatch conditions ( $t(58) = 2.3$ ,  $p < 0.05$ , Cohen's  $d = 0.59$ ), showing that participants looked longer to the opposite object in the mismatch condition. The main effects of Group and Condition did not reach significance, as well as the interaction between Group and Condition (all  $F_s < 0.001$ , all  $p_s > 0.99$ ). In order to assess the effect of clinical symptoms, as indicated by the AQ score, and participants' age, these two factors were introduced as covariates separately in two analyses of covariance (ANCOVA). The same main



**Figure 5.** Means of the baseline difference score on the four AOIs for the ASD and the NT groups. Error bars indicate the standard error of the mean (SEM).

AOIs: areas of interest; ASD: autism spectrum disorder; NT: neuro-typical.

effects and interactions reported above from the ANOVA remained significant, suggesting that no variance in relative looking time could be explained by the two covariates.

Because it is of particular relevance to our hypothesis, direct comparisons of relative looking time to the target object between groups were done for each condition separately, although the three-way interaction between AOI, Condition, and Group did not reach significance ( $F(6, 168)=1.68, p=0.19$ ). All  $p$ -values of these analyses were corrected using Holm's (1979) procedure. A significant difference in relative looking time to the target object between groups was found on the static ( $t(28)=3.8, p<0.005$ , Cohen's  $d=1.39$ ) and the mismatch conditions ( $t(28)=2.6, p<0.05$ , Cohen's  $d=0.95$ ), showing that the NT group looked longer to the target than the ASD group, but not for the familiarization condition ( $t(28)=1.7, p=0.1$ ).

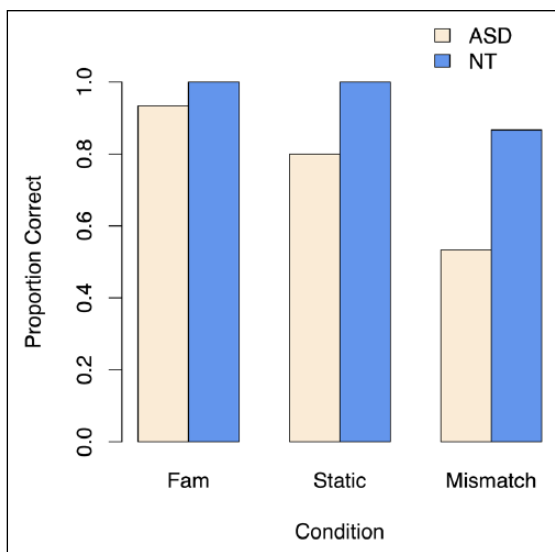
When the DS of relative looking time during baseline and response segments was analyzed (see Figure 5), a significant main effect of AOI was observed ( $F(3, 84)=76, p<0.001, \eta^2=0.56$ ). Further investigation of this effect by means of paired samples  $t$ -tests showed that participants' relative looking differed more between baseline and response segments to the target object compared to all other objects (all  $ps<0.001$ ). A significant interaction between AOI and condition was also found ( $F(6, 168)=11.6, p<0.001, \eta^2=0.18$ ). To investigate this interaction in detail, paired samples  $t$ -tests were used to compare the baseline DS for each AOI between conditions. These analyses showed that the baseline DSs of the target object were lower in the static and mismatch conditions compared to the familiarization condition (all  $ps<0.001$ ). Additionally, the baseline DSs of the distractors were

lower in the familiarization condition compared to the static and mismatch conditions (all  $ps<0.05$ ). The main effects of Group and Condition did not reach significance, as well as the interaction between Group and Condition and the interaction between Group, Condition, and AOI (all  $F_s<0.9$ , all  $ps>0.5$ ).

One-sample  $t$ -tests revealed that baseline DSs of the target object were significantly more than zero (all  $ps\leq 0.05$ ), suggesting that participants looked longer to the target in the response segment compared to the baseline. Although not all other differences were significant, the general trend showed that participants looked less to all other objects in the response segment compared to the baseline. The  $t$ -tests showed that, for the NT group, all baseline DSs in all three conditions for the two distractors and the opposite object were significantly less than zero (all  $ps<0.05$ ), except for one of the distractors in the static condition ( $t(14)=-1.35, p=0.2$ ). As for the ASD group, all baseline DSs in the familiarization condition for the two distractors and the opposite object were significantly less than zero (all  $ps<0.005$ ) and only for one of the distractors in the static and mismatch conditions (all  $ps<0.01$ ; all other  $ps>0.3$ ).

No significant correlations were observed between looking time to the face in the learning trials and relative looking time to the target during the response trials (all  $r_s<\pm 0.4, ps>0.1$ ). When the correlation between DS and relative looking time to target was assessed, a significant positive correlation was observed for the ASD group in the static condition ( $r=0.58, p<0.05$ ) and the mismatch condition ( $r=0.76, p<0.01$ ) and for the NT group in the static ( $r=0.62, p<0.02$ ) and the mismatch ( $r=0.86, p<0.001$ ) conditions.





**Figure 6.** Proportion of participants who selected the correct target explicitly after the presentation of each condition for both groups in each condition. ASD: autism spectrum disorder; NT: neuro-typical.

### Explicit responses

Chi-square tests showed a significant difference in the number of participants in the ASD group who selected the target compared with those in the NT group on the mismatch condition only ( $\chi^2(1, 28) = 5.79, p < 0.05$ ; all other  $ps > 0.2$ ). In Figure 6, the proportion of participants who selected the correct object is presented. Post hoc binomial tests showed that the proportion of participants who selected the target significantly differed from chance in all conditions (all  $ps < 0.05$ ; chance = 25%). When the chance level was set to 50%, given that only one of the three additional objects was a viable distractor, then the proportion of participants from the ASD group who selected the target did not differ from chance in the mismatch condition ( $p > 0.1$ ).

Assessment of the correlation between implicit and explicit responses revealed a significant positive correlation for the ASD group in the static condition ( $r = 0.73, p < 0.01$ ) and mismatch condition ( $r = 0.67, p < 0.01$ ). As for the NT group, a significant positive correlation between implicit and explicit responses was observed in the mismatch condition ( $r = 0.83, p < 0.001$ ). These correlations showed that participants who looked longer to the target object during the response segment were more likely to select that object in their explicit responses. All participants in the NT group had correct explicit responses in the static condition; therefore, no correlation test was possible. When correlations between looking time to the face in the learning trials and the explicit responses were assessed, a significant positive correlation was only observed in the ASD group in the static condition ( $r = 0.58, p < 0.05$ ).

### Demographic data

No significant correlations were observed between any of the additional measures and their relative looking time to the target during the response segment in either group (all  $rs < \pm 0.43$ , all  $ps > 0.1$ ).

### Discussion

In this study, we investigated whether adults with autism can learn novel words by referring to the direction of gaze. Moreover, we were interested in examining whether adults with autism prioritize gaze cues (i.e. a social cue) over a nonsocial cue, when they were presented simultaneously but in conflict with each other. To this end, participants observed on an eye-tracking screen an animated actress, presented with two novel objects in front of her. The actress looked at one of the objects (the target) and labeled it with a novel name, while completely ignoring the other object (the opposite). In the static condition, the opposite object was stationary throughout the learning trial. In the mismatch condition, the opposite object was cued by a nonsocial cue (i.e. jiggling), while the target was cued by the social cue (i.e. gaze direction). Results of participants' looking times and their explicit responses in the static condition show that adults with ASD, as well as the NT, were able to choose the correct referent of the novel word indicating that they relied on the actress' gaze cue during the learning trial. In contrast, results of the mismatch condition show that the performance of the ASD group dropped to chance level, while the NT group choose the correct referent almost as well as they did in the static condition. We interpret these findings as evidence that adults with autism have some understanding of the referential nature of others' gaze, but not to the same extent as NT adults.

Whereas previous studies demonstrated that children with autism have difficulties in relying on gaze cues in word learning (e.g. Akechi et al., 2011; Preissler and Carey, 2005), this study demonstrated that adults with autism are able to do so. This parallels results of the ToM literature where it has been reported that people with ASD develop the ability to solve tasks that require attribution of mental states later than their typically developing peers (Happé, 1995). These findings suggest that people with autism develop some of the social-cognitive competencies that are characteristic for typically developing people. Yet, this development seems to be more effortful, and consequently, these competencies do appear later.

How can the discrepancy with the other findings then be explained? First, one could argue that our employment of animated drawings supported participants' learning. By using these highly controlled stimuli, we were able to control for any unnecessary distractions, which might be a problem when using life stimuli. ASD participants might be overwhelmed by such irrelevant elements in a scene

and might focus a lot of their attention on it in a life situation (see Falck-Ytter and Von Hofsten, 2010), and therefore, their competence to use social cues might be underestimated. However, this explanation is unlikely given that also other studies relied on animated agents, but nevertheless demonstrated problems in gaze understanding in people with autism (e.g. Böckler et al., 2014).

Second, this study examined ASD adults, whereas the previous studies mostly focused on ASD children. It is possible that people with ASD develop compensatory mechanisms to overcome the problem of using social cues to direct their behavior (Elsabbagh and Johnson, 2010). This interpretation is supported by the following happening during one of the test sessions. One participant from the ASD group reported that, at the beginning of the experimental session, he was not paying attention to the face of the actress, and it took him conscious effort to attend to the actress' face to see where she was looking. After examining that participant's gaze pattern, it turned out that he looked significantly longer to the target object in the static condition, and he chose it from the set of cards, demonstrating that this strategy might have helped him in choosing the correct referent of the novel word. This suggests that people with autism might acquire reflexive compensatory strategies in the course of development, which help them to overcome their initial problem in appreciating the referential nature of other's social cues.

Interestingly, a different pattern of results was found when the distractor's saliency was increased during the labeling action (i.e. the mismatch condition), that is, when a saliency cue interfered with the concurrent social cue given by the actress. Here, the performance of the ASD group dropped to chance level, while the NT group was still able to choose the correct referent of the novel word. Even in their relative looking time, the ASD group looked significantly less to the target, relative to the NT group. Yet, participants from both groups showed a significant increase in relative looking to the target object after its name was mentioned relative to the baseline segment in the mismatch condition. This suggests that, despite the ASD group's ability to distinguish the correct referent for the novel word in the mismatch condition, they still choose the incorrect object almost half of the time.

How can this impaired performance of the ASD group in the mismatch condition be explained? We offer two explanations. First, participants need to disengage their attention from the salient object and reallocate it to the target (Gliga et al., 2012). However, individuals with ASD were described to have problems in disengaging their attention from an object (Landry and Bryson, 2004) and in inhibiting distractors (Adams and Jarrold, 2012). Following this line of argumentation, one could say that, in our mismatch condition, the ASD group was not able to ignore the opposite object during the learning trials. This might have led to attributing the novel word to the opposite object.

However, our results cannot be explained exclusively by this account. First, analyses of looking times to the face of the actress showed no difference between groups or conditions. Yet, we would have expected a decrease in looking time to the actress for the ASD group in the mismatch condition, if they would have had problems in disengaging from the salient distractor. Second, the analysis of the DS revealed that the ASD group looked for the same duration to both the target and the opposite objects, indicating that they processed these objects to the same extent.

A second explanation is the social-cognitive account, which suggests that, although ASD participants can in principle use social cues, they do not prefer them to other (nonsocial) cues. In other words, it is possible that in the mismatch condition of this study, both cues were valid to the same extent for adults with ASD. This might have led to confusion as to which cue should they follow. We know from previous studies that children with ASD can—similarly to typically developing children (Hollich et al., 2000; Houston-Price et al., 2006)—rely on saliency cues alone to choose the correct referent of a novel word, even without the presentation of a matching social cue (Luyster and Lord, 2009). Likewise, Akechi et al. (2011) have shown that children with ASD benefit from the presence of a matching saliency cue with the gaze cue to learn a new word-object association. The DS in our results also supports this hypothesis because the ASD group did not prefer the salient object during the labeling segment, but they looked at both the target (cued by the social cue) and the opposite (cued by the nonsocial cue) for roughly the same amount of time, showing that both objects were of the same relevance to them and they were not able to distinguish which one was actually the target. It is worth noting that our paradigm assessed whether adults with ASD can rely on gaze cues to guide their word learning spontaneously. Future studies need to clarify whether the performance of people with ASD would be intact in more explicit situations of word learning in the presence of conflicting cues.

In addition to the results discussed above, we found group differences in relative looking time to the target object during response trials in both the static and the mismatch conditions, showing that the ASD group's relative looking to the target was lower than that of the NT group. In the static condition, where both groups looked more to the target than to the other objects, this could be explained by the speed at which people with ASD process visual stimuli. Faster reaction times to visual stimuli were reported for people with ASD compared to an NT control group (Chawarska et al., 2003). This suggests that the ASD group might have been faster in checking relevant items in the environment than the NT group, after which they started investigating the rest of the scene.

Interestingly, we also found a positive correlation between DS and task performance. This indicates that the

more the ASD participants were able to prioritize the gaze over the saliency cue, the more they preferred looking at the target after its name was mentioned during the response segment. This finding suggests that participants' test performance indeed measured their reliance on the gaze cue during the learning phase. It also points to individual differences within the ASD group, suggesting that some were able to rely on the gaze cue even in the mismatch condition, while overall group performance was not as good as the NT group. Further research is necessary to explore individual differences in social-cognitive abilities in people with autism.

It should be noted that the implicit and the explicit measures provided converging results. Moreover, the positive correlations between participants' looking behavior in the test trials and their explicit responses in each condition indicate that both measures assessed the same ability, strengthening the validity of our task. This relation is important for two further reasons. First, by demonstrating the effect across two different response modalities, we show that preferential looking paradigms can be a valid tool to assess social-cognitive abilities in general. Second, it suggests that participants did not merely learn an association between an utterance and an object (a possible objection to implicit measures of word learning), but that they indeed acquired a novel word (see Bannard and Tomasello, 2012).

In conclusion, this study is the first to demonstrate that adults with ASD are fully capable of spontaneously using the gaze of another person to select the correct referent of a novel word. Yet, when there is a conflicting saliency cue with the gaze cue, the performance of the NT group remained intact, while that of the ASD group dropped to chance level. This puts forward a proof that gaze understanding develops in people with ASD, however not to the same extent as their typically developing peers.

### Acknowledgements

We thank all participants who took part in this study. We are grateful to Nicosia Nieß and Gertrud Niggemann (Autismus Oberbayern e.V.), Martina Schabert (Autismuszentrum Oberbayern), and Martin Sobanski (Heckscher-Klinikum gGmbH) for their support. We also thank Tabea Schädel, Veronika Sophie Eisenschmid, and Verena Rampeltshammer for their help with data acquisition and Samia Saade for her help with preparing the stimuli. We finally thank our reviewers for their helpful comments.

### Funding

This research was funded by a grant from the Volkswagen Foundation (Research group "Knowledge through interaction," grant number Az. 86 755).

### References

Adams NC and Jarrold C (2012) Inhibition in autism: children with autism have difficulty inhibiting irrelevant distractors but not

prepotent responses. *Journal of Autism and Developmental Disorders* 42(6): 1052–1063.

Akechi H, Senju A, Kikuchi Y, et al. (2011) Do children with ASD use referential gaze to learn the name of an object? An eye-tracking study. *Research in Autism Spectrum Disorders* 5(3): 1230–1242.

Baldwin DA (1991) Infants' contribution to the achievement of joint reference. *Child Development* 62(5): 875–890.

Baldwin DA (1993) Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language* 20: 395–418.

Bannard C and Tomasello M (2012) Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds? *PLoS ONE* 7(11): 1–7.

Baron-Cohen S, Baldwin DA and Crowson M (1997) Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development* 68(1): 48–57.

Baron-Cohen S, Leslie AM and Frith U (1985) Does the autistic child have a "theory of mind?" *Cognition* 21(1): 37–46.

Baron-Cohen S, Wheelwright S, Skinner R, et al. (2001) The autism-spectrum quotient (AQ): evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31(1): 5–17.

Bergmann C, Paulus M and Fikkert P (2012) Preschoolers' comprehension of pronouns and reflexives: the impact of the task. *Journal of Child Language* 39: 777–803.

Böckler A, Timmermans B, Sebanz N, et al. (2014) Effects of observing eye contact on gaze following in high-functioning autism. *Journal of Autism and Developmental Disorders* 44(7): 1651–1658.

Bui R, Buliung RN and Rummel TK (2012) aspace: a collection of functions for estimating centrographic statistics and computational geometries for spatial point patterns. R package version 3.2. Available at: <http://CRAN.R-project.org/package=aspace>

Chawarska K, Klin A and Volkmar F (2003) Automatic attention cueing through eye movement in 2-year-old children with autism. *Child Development* 74(4): 1108–1122.

Elsabbagh M and Johnson MH (2010) Getting answers from babies about autism. *Trends in Cognitive Sciences* 14(2): 81–87.

Falck-Ytter T and Von Hofsten C (2010) How special is social looking in ASD: a review. *Progress in Brain Research* 189: 209–222.

Freitag CM, Retz-Junginger P, Retz W, et al. (2007) Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ)–die Kurzversion. *Zeitschrift für Klinische Psychologie und Psychotherapie* 36(4): 280–289.

Frith CD and Frith U (2008) Implicit and explicit processes in social cognition. *Neuron* 60: 503–510.

Frith CD and Frith U (2012) Mechanisms of social cognition. *Annual Review of Psychology* 63: 287–313.

Gillespie-Lynch K, Elias R, Escudero P, et al. (2013) Atypical gaze following in autism: a comparison of three potential mechanisms. *Journal of Autism and Developmental Disorders* 43(12): 2779–2792.

Gluga T, Elsabbagh M, Hudry K, et al. (2012) Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development* 83(3): 926–938.

- Greene DJ, Colich N, Iacoboni M, et al. (2011) Atypical neural networks for social orienting in autism spectrum disorders. *NeuroImage* 56(1): 354–362.
- Grimm H (2001) *Sprachentwicklungstest für drei-bis fünfjährige Kinder: Diagnose von Sprachverarbeitungsstörungen und auditiven Gedächtnisleistungen*. Göttingen: Hogrefe-Verlag GmbH & Co. KG.
- Happé F (1995) The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development* 66(3): 843–855.
- Hollich GJ, Hirsh Pask K, Golinkoff RM, et al. (2000) Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development* 65(3): i–vi, 1–135.
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6: 65–70.
- Houston-Price C, Plunkett K and Duffy H (2006) The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology* 95(1): 27–55.
- Kuhn G, Benson V, Fletcher-Watson S, et al. (2010) Eye movements affirm: automatic overt gaze and arrow cueing for typical adults and adults with autism spectrum disorder. *Experimental Brain Research* 201(2): 155–165.
- Landry R and Bryson SE (2004) Impaired disengagement of attention in young children with autism. *Journal of Child Psychology and Psychiatry* 45(6): 1115–1122.
- Lawrence MA (2013) ez: easy analysis and visualization of factorial experiments. R package version 4.2-2. Available at: <http://CRAN.R-project.org/package=ez>
- Lehrl S (2005) *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. unveränderte Aufl. Balingen: Spitta Verlag, 5.
- Luyster R and Lord C (2009) Word learning in children with autism spectrum disorders. *Developmental Psychology* 45(6): 1774–1786.
- Moore C, Angelopoulos M and Bennett P (1999) Word learning in the context of referential and salience cues. *Developmental Psychology* 35(1): 60–68.
- Paulus M (2011) How infants relate looker and object: evidence for a perceptual learning account on gaze following in infancy. *Developmental Science* 14: 1301–1310.
- Paulus M and Fikkert P (2014) Conflicting social cues: fourteen- and 24-month-old infants' reliance on gaze and pointing cues in word learning. *Journal of Cognition and Development* 15(1): 43–59.
- Preissler MA and Carey S (2005) The role of inferences about referential intent in word learning: evidence from autism. *Cognition* 97(1): B13–B23.
- R Core Team (2013) *R: A Language and Environment for Statistical Computing*, version 3.0.2. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>
- Salvucci DD and Goldberg JH (2000) Identifying fixations and saccades in eye-tracking protocols. In: *Proceedings of the 2000 symposium on eye tracking research and applications, (ETRA '00)*, Palm Beach Gardens, FL, USA, 6–8 November, pp. 71–78. New York: ACM.
- Schwarzkopf S, Schilbach L, Vogeley K, et al. (2014) “Making it explicit” makes a difference: evidence for a dissociation of spontaneous and intentional level 1 perspective taking in high-functioning autism. *Cognition* 131: 345–354.
- Senju A, Southgate V, Miura Y, et al. (2010) Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology* 22(2): 353–360.
- Senju A, Southgate V, White S, et al. (2009) Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science* 325: 883–885.
- Senju A, Tojo Y, Dairoku H, et al. (2004) Reflexive orienting in response to eye gaze and an arrow in children with and without autism. *Journal of Child Psychology and Psychiatry* 45(3): 445–458.
- Tomasello M (2006) Social-cognitive basis of language development. In: Brown K (ed.) *Encyclopedia of Language and Linguistics*. 2nd ed. Oxford: Elsevier, 459–462.
- Waxman SR and Gelman SA (2009) Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences* 13(6): 258–263.
- Weiss RH (2006) *Grundintelligenztest Skala 2–Revision (CFT 20-R)*. Göttingen: Hogrefe.
- Wickham H (2007) Reshaping data with the reshape package. *Journal of Statistical Software* 21(12): 1–20. Available at: <http://www.jstatsoft.org/v21/i12/>
- Wu R and Kirkham NZ (2010) No two cues are alike: depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology* 107(2): 118–136.
- Zeileis A and Grothendieck G (2005) zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software* 14(6): 1–27. Available at: <http://www.jstatsoft.org/v14/i06/>