AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001040

Audio-Visual Integration in Nonverbal or Minimally Verbal Young Autistic Children

Mikhail Kissine[1, 2, 3], Julie Bertels[3, 4], Nicolas Deconinck[1, 3, 5], Gianfranco Passeri[5], and Gaétane Deliens[1, 3, 6]

[1] ACTE, Université libre de Bruxelles
[2] LaDisco, Université libre de Bruxelles
[3] ULB Neuroscience Institute, Université libre de Bruxelles
[4] ULBabyLab at CRCN, Université libre de Bruxelles
[5] Centre de Référence Autisme "Autrement," Hôpital Universitaire Reine Fabiola, Brussels, Belgium
[6] CRCN, Université libre de Bruxelles

Low integration of speech sounds with the mouth movements likely contributes to language acquisition disabilities that frequently characterize young autistic children. However, the existing empirical evidence either relies on complex verbal instructions or merely focuses on preferential gaze on in-synch videos. The former method is clearly unadapted for young, minimally, or nonverbal autistic children, while the latter has several biases, making it difficult to interpret the data. We designed a Reinforced Preferential Gaze paradigm that allows to test multimodal integration in young, nonverbal autistic children and overcomes several of the methodological challenges faced by previous studies. We show that autistic children have difficulties in temporally binding the speech signal with the corresponding articulatory gestures. A condition with structurally similar nonsocial video stimuli suggests that atypical multimodal integration in autism is not limited to speech stimuli.

Keywords: autism, language delay, audio-visual integration, eye-tracking

Supplemental materials: https://doi.org/10.1037/xge0001040.supp

Autism spectrum disorder (ASD) is a lifelong neurobiological developmental condition, whose prevalence is currently estimated at more than one child over 70 (Autism and Developmental Disabilities Monitoring Network, 2016). ASD core behavioral features crystallize as, on the one hand, marked difficulties in verbal and nonverbal communication and social interaction, and, on the other hand, a restricted repertoire of repetitive or stereotypical interests and behaviors (American Psychiatric Association, 2013). Although difficulties in the sociopragmatic aspects of language use characterize all stages of the cognitive and linguistic development of autistic individuals (e.g., Deliens et al., 2018; Kim et al., 2014), language deficits are currently considered as a specifier of an autism diagnosis (American Psychiatric Association, 2013).[1] However, language acquisition is considerably delayed in around 60% of autistic children, and around 30% of them do not develop functional language at all (e.g., Anderson et al., 2007; Baghdadli et al., 2018; Kim et al., 2014; Wodka et al., 2013).

While language is a crucial factor for outcomes in autism, little is known about the causes of language disabilities in children on the autism spectrum. One prominent hypothesis is that lack of sensitivity to social stimuli—and more particularly to facial cues—in the early stages of life has a cascading effect on the acquisition of language and on communication skills. Some retrospective analyses suggest that lower social impairment or better joint attention skills correlate with language levels in ASD (Wodka et al., 2013; Yoder et al., 2015). However, in other large longitudinal or prospective studies, sociocommunicative variables do not systematically predict language outcomes, especially once nonverbal IQ is factored in (Anderson et al., 2007; Bennett et al., 2015; Ellis Weismer & Kover, 2015; Thurm et al., 2015). Several authors recently hypothesized that sociocommunicative disabilities in autism co-occur with or are partly caused by a domain-general atypical development in perceptual processing (Bahrick & Todd, 2012;

Mikhail Kissine https://orcid.org/0000-0001-6693-9772

First and foremost, we warmly thank all the children who took part in our study, as well as their families and our partner institutions. We are also extremely grateful to the Jean-François Peterbroeck Foundation, the Support Fund Marguerite-Marie Delacroix, and the Foundation ULB for their generous support of the ACTE lab. Mikhail Kissine is a 2019–2022 Francqui Research Professor.

Correspondence concerning this article should be addressed to Mikhail Kissine, ACTE, Université libre de Bruxelles, CP 175, Avenue F.D. Roosevelt, 1050 Bruxelles, Belgium. Email: Mikhail.Kissine@ulb.be

[1] Following the preferences increasingly expressed by the autism community (Kenny et al., 2016), throughout the article, we will use "autistic child (or individual)" rather than "child (or individual) with autism" or "child (or individual) with autism spectrum disorder."

Feldman et al., 2018; Robertson & Baron-Cohen, 2017; Stevenson et al., 2016).

Interestingly, there is growing evidence that abnormalities in social orientation are not present from birth, but rather gradually emerge during the first year of life (Jones & Klin, 2013), which suggests that an inherent processing dysfunction prevents autistic infants from reaching early milestones in language development. In typically developing (TD) infants, the period between 6 and 12 months corresponds to the emergence of sensitivity toward native phonological categories, such as, for instance, the/r/–/l/ contrast for American but not Japanese babies (e.g., Vihman, 2014). Multimodal sensory integration is central to early child development (Bahrick et al., 2004). In typical development, integrating visual articulatory cues with speech signal plays an important role in the acquisition of the sounds of one's native language (Kuhl & Meltzoff, 1984; Teinonen et al., 2008), as dynamic changes in the shape of the mouth are correlated with a number of salient acoustic aspects of speech, thus adding redundant visual information to the auditory input (Chandrasekaran et al., 2009). Accordingly, TD infants spend an increased amount of time looking at the speaker's mouth between 6 and 12 months—the crucial period for the emergence of one's native tongue phonological categories—but redirect the focus on the eye region after that period (Lewkowicz & Hansen-Tift, 2012).

There are good reasons to hypothesize that young autistic children fail to map mouth movement on speech sounds, thus missing a crucial bootstrap for accessing language. Older autistic children and adults are significantly less prone than their TD peers to use visual information in order to improve perception of speech in noise, have a lower ability to read silent speech from lip movements, and are less influenced by mismatching visual information when hearing speech in McGurk-type paradigms (e.g., de Gelder et al., 1991; Irwin et al., 2011; Irwin & Brancazio, 2014; Mongillo et al., 2008; Smith & Bennetto, 2007; Stevenson et al., 2016, 2018). In the same vein, Magnée et al. (2008) report reduced ERP correlates for audio-visual integration in adults with a diagnosis of pervasive developmental disorder. Furthermore, autistic toddlers who attend to the speaker's mouth are more likely to develop language than those who do not (Campbell et al., 2014).

However, most of the existing studies on audio-visual integration in autism rely on verbal instructions or explicit behavioral responses, and—partly for this reason—focus on verbal autistic children and adults. In that respect, literature on multimodal integration in ASD suffers from the same sampling bias as the rest of psycholinguistic research in autism, which tends to dramatically overresearch verbal individuals. Collecting rigorous behavioral evidence on younger, non or minimally verbal autistic children is rife with practical and methodological challenges (Tager-Flusberg et al., 2017). Yet, if language development in ASD is impacted by low audio-articulatory integration, it is of paramount importance to assess this processing component precisely in children who experience language acquisition delay. In older, verbal autistic children, language acquisition and cognitive development may have entailed an improvement in multimodal integration, thus obfuscating potential group differences that could have been manifest at earlier developmental stages, closer to the onset of language delay (see, also Bahrick & Todd, 2012).

In typical development, the crucial evidence for precocious multimodal integration in speech comes from studies that employ the preferential gaze paradigm. TD infants display sensitivity to audio-visual asynchrony in native speech by preferentially gazing toward the recordings of a speaking face whose articulatory movements match the simultaneously played audio recording versus one with a mouth movement-audio mismatch (e.g., Hillairet de Boisferon et al., 2017; Kuhl & Meltzoff, 1984; Patterson & Werker, 2003). This is also the method used in two studies that investigated multimodal integration in young (around 5 years of age) autistic children (Bebko et al., 2006; Righi et al., 2018). Bebko et al. (2006) report that autistic children displayed no preferential gaze toward the in-synch videos, with recordings of a woman face telling a story or simply counting forward. Likewise, Righi et al. (2018) found no preference toward in-synch, rather out-of-synch (by .3 s, .6 s, or 1 s) recordings of a speaking woman's face.

These two studies represent a decisive step toward a better understanding of multimodal integration in ASD. However, in some respects this evidence is somewhat difficult to interpret. The implicit rationale behind using preferential gaze paradigms in this context is that lower audio-visual integration in autistic versus TD children would surface as a group difference in fixation distributions between in- and out-of-synch stimuli. Accordingly, lack of preference for any type of stimuli in autistic children is interpreted as reflecting a difficulty in distinguishing between in- and out-of-synch audio-visual alignment. As mentioned earlier, there is robust evidence that TD infants display a preference toward in- versus out-of-synch audio-visual alignment (e.g., Hillairet de Boisferon et al., 2017; Kuhl & Meltzoff, 1984; Patterson & Werker, 2003). However, it is unclear whether one can safely presuppose that in older TD toddlers or preschoolers temporal alignment of the audio and video signals would lead to a preference for the in-synch recording. In fact, in Bebko et al. (2006) TD children displayed no preference for the in-synch simple linguistic stimuli and only a weak one for more complex ones, so that no significant difference emerged with the ASD group. Likewise, in Righi et al. (2018) TD children preferentially gazed toward the in-synch video only when the other one was out-of-synch by .6 s or 1 s, but not when the temporal delay was of .3 s. These authors also found no difference in gaze allocation between their ASD and TD groups with .3 s, .6 s, or 1 s out-of-synch videos.

Lack of preferential gaze on synchronous videos may also be caused by increased fixations toward the *a*synchronous side, because the unusual misalignment attracts children's gaze (or, albeit this is less likely, due to an inherent preference for asynchrony). As an illustration of the ambiguity inherent in the interpretation of preferential gaze data, consider the article by Guiraud et al. (2012). These authors also report a difference in speech audio-visual integration in infants with low- and high-likelihood to receive a diagnosis of ASD. However, contrary to Bebko et al. (2006) and Righi et al. (2018), their interpretation is based on the fact that low-likelihood infants fixated *more* the speaker's mouth when the video recording was incongruent with the acoustic signal. Previous studies that compare visual exploration of in- versus out-of-synch videos fail to distinguish between two conflicting causal hypotheses about children's visual attention: On the one hand, children may prefer to look at familiar, in-synch signals, but, on the other hand, their gaze can also be strongly attracted by the unusual out-of-synch video. In other words, it is unclear how the

familiarity of in-synch temporal alignment competes with the novelty of out-of-synch stimuli.

Moreover, it is possible that children's visual attention is driven by salient changes in the stimuli—articulatory movements in the case of speech stimuli—independently of audio-visual temporal alignment. Examining fixation curves (viz., gaze trajectories over time) may help determine whether factors other than familiar, in-synch audio-visual alignment attract children's visual attention. In studies that use preferential gaze paradigms to investigate audio-visual integration in ASD, the salient mouth movements, by definition, occur at different time points in the two simultaneously presented videos. Therefore, if mouth movements on each video attract children's attention, the curves of fixations on the in-synch and out-of-synch videos should display a periodic alternation, corresponding to the occurrence of salient articulatory events.

Unfortunately, in both Bebko et al. (2006) and Righi et al. (2018), experimental trials lasted around 13 s, and both studies reported gaze data as proportional fixation averages per trial, which makes it impossible to assess detailed temporal trajectories of eye-fixations. Furthermore, some children may devote a significant portion of such a long stretch of time to visually explore facial regions other than the mouth. Because these regions are less informative as to the temporal alignment with the acoustic signal, they may be explored indifferently in the in- or out-of-synch video. Given the atypical patterns of facial exploration documented in ASD, it is possible that autistic children spend less time looking at mouth altogether, thus having less opportunity to detect audio-visual (a)synchrony. And Righi et al. (2018) do report overall less fixations on the mouth in their ASD group. (That said, other studies found increased attention to the mouth region in autism; e.g., Klin et al., 2002).

The present study builds on previous research, but aims at avoiding the methodological issues just discussed, in order to reach a better assessment of young, minimally verbal autistic children's capacity to integrate mouth movements with the speech signal. On the top of the mouth region, head and face movement may contribute to parsing the speech stream (Munhall et al., 2004). However, because visual exploration of faces is known to be atypical in autistic children (Jones & Klin, 2013), using full face recordings to test audio-visual integration of speech signals may introduce further biases. In order to focus children's attention on visible articulatory gestures, the video stimuli of our *speech condition*—we will turn to our other condition in a short while—are limited to the mouth region, with the rest of the face being masked. Furthermore, each stimulus consists in a 5 s recording of three identical consonant-vowel syllables, so that three clear articulatory movements can be easily mapped on three salient acoustic events, associated with the consonant. As just argued, mere comparison of gaze distributions between in- or out-of-synch stimuli may not be sufficiently informative as to the children's multimodal integration skills. In order to circumvent this issue we implemented a reinforcement-based anticipation method. This paradigm rests on the same conditioning mechanism that underlies the "visually reinforced infant speech discrimination" paradigms, widely used in the developmental literature (e.g., Werker & Tees, 1984). In these classic paradigms, children are conditioned to associate an attractive reinforcement with stimuli belonging to one category, but not with those belonging to another one. Once this association is operational, anticipation of the reinforcement may serve as a proxy for the participant's ability to categorize the target stimulus. Our objective in the current study is to determine the extent to which (autistic) participants distinguish between in- and out-of-synch stimuli. The logic underlying our paradigm is thus that those participants who can distinguish between in- and out-of-synch stimuli can be conditioned to associate a reinforcement with one of these two types of stimuli, whereas in those who struggle to distinguish between in- and out-of-synch stimuli, the reinforcement should be less operational. Relying on such implicit reinforcement mechanisms in researching autistic children is warranted by the fact that current evidence indicates that associative, implicit learning is not affected in autism (see the meta-analyses in Foti et al., 2015; Obeid et al., 2016). In each of our trials, the *stimulus presentation phase* is followed by a 1 s *transition* blank screen, after which starts a 3 s *reinforcement phase*. In the speech condition, reinforcements consist in different visually attractive animations, superimposed on the last frame of the corresponding video. The position of the reinforcement can be anticipated only based on temporal alignment between the video and the audio components of the stimuli: for half of the children in each group (TD or ASD), the reinforcement consistently appeared on the side of the in-synch video (*synchronous version*) and, for the other half, the reinforcement consistently appeared on the side of the out-of-synch video (*asynchronous version*). Consequently, anticipative gaze toward the location of the reinforcement during the transition phase—to the side of the in-synch video in the synchronous version and to the side of the out-synch video in the asynchronous version—is indicative of the capacity to temporally bind the acoustic and the video signals. In our reinforcement paradigm the stimulus and transition phases are analyzed separately, which allows to minimize the confounding influence novelty or familiarity effects may exert in a passive viewing setting. Because TD children are highly sensitive to temporal asynchrony between voice and mouth, we expect them to anticipatively gaze toward the reinforcement in the transition phase of the speech condition, based on the reinforced type of video stimuli (in- or out-of-synch, depending on the version to which they are assigned). In the ASD group, by contrast, poor integration of speech and articulatory information should compromise the distinction between in- and out-of-synch video stimuli. As a consequence, autistic children are expected to display significantly lower anticipation of the reinforcement in the transition phase.

Another burning research issue is the causality of deficient multimodal integration in ASD. Relationship between sensory processing and higher-order cognitive symptoms of autism are often modeled either in a "top-down" or a "bottom-up" fashion (Robertson & Baron-Cohen, 2017). One straightforward top-down hypothesis could be that lack of multimodal integration in language arises because autistic children do not pay sufficient attention to faces (and particularly to the mouth region). Under such a view, the atypical developmental trajectory that becomes evident toward the end of the first year of life would be a consequence of a lack of interest in social information, inherent in ASD. This explanation would be consistent with models of autism centered on an innate deficit in social motivation and skills (Chevallier et al., 2012). A contrasting, or rather complementary, bottom-up approach is to posit that a lower-level, atypical sensory integration style prevents autistic infants from mapping articulatory cues, gathered from observing mouth movements, on acoustic information. In other words, it is possible that low capacity to integrate visual and

acoustic information contributes to low attention to the mouth region in autistic infants. Not only would then these infants miss the first crucial steps in language acquisition, they would also fail to be reinforced to visually explore faces as sources of valuable information. This alternative hypothesis, which links language disabilities in autism to an atypical lower-level sensory processing style (Robertson & Baron-Cohen, 2017), is consistent with the mounting evidence for a disruption in temporal binding of audio and visual information in ASD (e.g., Bahrick & Todd, 2012; Foss-Feig et al., 2010; Stevenson et al., 2014; Turi et al., 2016), and, more generally, with models of atypical sensory processing in autism (e.g., Mottron et al., 2006; Pellicano & Burr, 2012).

A relatively straightforward way to address this research question is to compare audio-visual integration between speech and nonsocial stimuli. Lower integration in speech and in nonsocial conditions alike by the ASD group would constitute evidence for a domain-independent deficit in multimodal temporal binding. Another possibility is that, in autistic children, lower audio-visual integration in speech combines with a TD-like performance when exposed to nonsocial stimuli. On the one hand, such a result could be interpreted as supporting the idea of an atypical processing of specifically social information; on the other hand, it is possible that speech events contain faster and more numerous changes, thus placing higher demands on multisensory processing, independently of their social nature (see Bahrick & Todd, 2012). As a step toward avoiding this ambiguity, it is crucial to match the frequency of changes, as we do below, between speech and nonsocial stimuli.

Again, most of the existing literature on nonsocial audio-visual integration in ASD focuses on highly verbal school-age children, teenagers, or adults and employs tasks that require explicit instructions and verbal responses. Results from this literature are somewhat mixed. For instance, Irwin et al. (2011) found that autistic children do not differ from TD children in detecting asynchrony between sine waves shapes and consonant-vowel syllables. Likewise, Stevenson et al. (2018) did not find group differences in the detection of asynchrony between flashes and beeps. However, Turi et al. (2016) report reduced recalibration to audio-visual misalignment in autistic adults, while Foss-Feig et al. (2010) found that the temporal interval during which audio stimuli (number of beeps) are likely to influence visual perception (number of flashes) is wider in autistic children than in their TD peers.

The study by Bebko et al. (2006) on 5-year-old autistic children, already discussed above, also included a nonsocial trial. Interestingly, unlike with linguistic stimuli, autistic children seemed to display a preference for the synchronous nonsocial videos, to the same extent as the TD group. This result suggests that these autistic children are sensitive to the audio-visual alignment of nonsocial events. However, there are several reasons for further exploring this finding. First, the ASD and TD groups in this study were of a relatively small size of 16 children per clinical group. Furthermore, children were exposed to a single nonsocial stimulus, which consisted in a 12 s video recording of a child "Mousetrap" game, where a ball follows a complex trajectory across pipes and various obstacles. This is a quite complex event, with many features that may influence gaze distribution independently of audio-visual (mis)alignment.

Klin et al. (2009) exposed autistic 2-year-olds, on one side of a computer screen, to point-light animations that corresponded to the motion executed during child games, such as peekaboo, coupled with the corresponding audio-recording; on the other side of the screen, the same animation was presented upside-down and played backward. Unlike children in TD and in developmental delay comparison groups, autistic children displayed no general preference toward the upward animation. However, when animations contained salient features that allowed an easy mapping on synchronous audio events, autistic children did preferably gaze at the upward video. This finding is interpreted by the authors as indicating a preference for nonsocial contingencies, over biological motion, in ASD. It also strongly suggests that young autistic children are sensitive to salient audio-visual synchrony. However, the essential aspect of the stimuli used by Klin et al. (2009) was biological motion, which was determining for higher preferential gaze in the comparison groups. For this reason, their study does not contain data that allows to truly compare audio-visual integration in nonsocial stimuli between ASD and TD groups.

In order to rigorously compare audio-visual integration between speech and nonsocial stimuli, we designed a *nonsocial condition*, structurally identical to the speech condition described above. Nonsocial stimuli had exactly the same structure as the 5 s speech videos, but consisted of animated cartoons of periodic object movements occurring three times in a row (e.g., a basketball bouncing on the ground, or drops of water leaking from a faucet), accompanied by the corresponding salient sound. In the nonsocial condition the reinforcements consisted in different amusing sequels to the corresponding video.

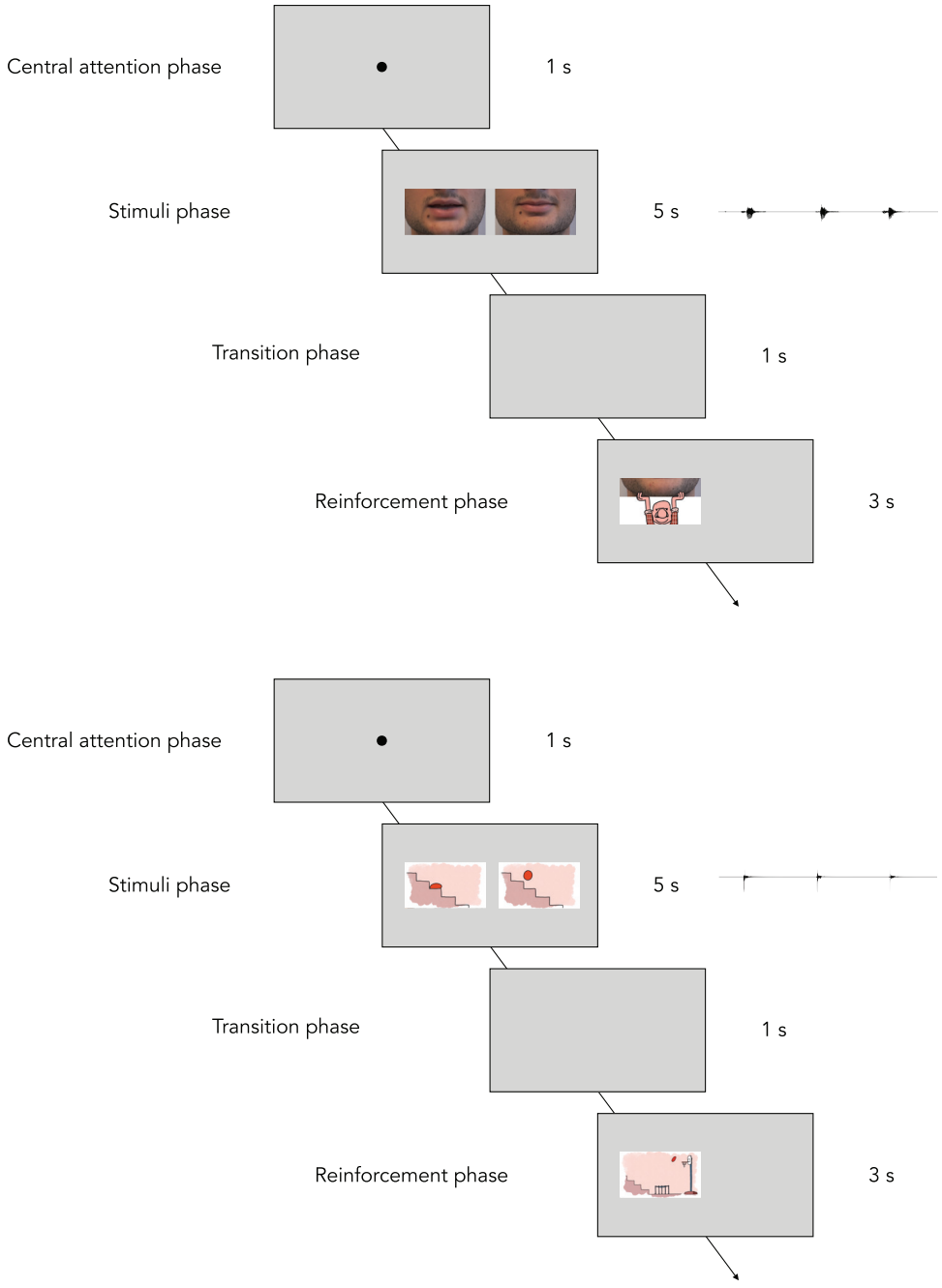
Method

Material and the Task

A set of 30 5-s long stimulus videos and their matching 3-s long reinforcements was created for each condition. Speech condition stimuli consisted in recordings of a woman's ($N = 15$) or a man's ($N = 15$) mouth, uttering the same occlusive consonant-vowel syllable three times in a row (e.g., */pupupu/*) with a short silence between each of the three repetitions (mean silence duration: 1.27 ± .31 s; mean syllable duration: .37 ± .11 s). Fifteen different syllables were assigned, each, to a different woman and 15 different syllables were assigned, each, to a different man (i.e., 15 different male and 15 different female speakers were used). The 30 reinforcement videos consisted in a visually attractive animation superimposed on the last frame of its matched stimulus video (e.g., a gardener trimming the beard around the mouth presented on the screen with a lawn mower). Nonsocial condition stimuli consisted in animated cartoons of periodic object movements occurring three times in a row (e.g., a basketball bouncing on the ground, or drops of water leaking from a faucet). Each movement was accompanied by an audio recording of the corresponding sound. The reinforcement consisted in an amusing sequel to its matched stimulus video (e.g., the basketball falling on the stairs, bouncing on a trampoline and coming to rest in a basketball hoop). Each trial began with a central attention phase in which a fixation dot was displayed in the center of the screen for 1 s, followed by a 5-s stimuli phase, a 1-s transition phase, and a 3-s reinforcement phase. See Figure 1 for an illustration of a trial course in each condition.

Animations presented in the reinforcement videos, as well as the stimulus videos for the nonsocial condition, were created by a professional illustrator using TV Paint, Adobe AfterEffects, and

Figure 1
Complete Trial Course: Speech (Top) and Nonsocial (Bottom) Conditions

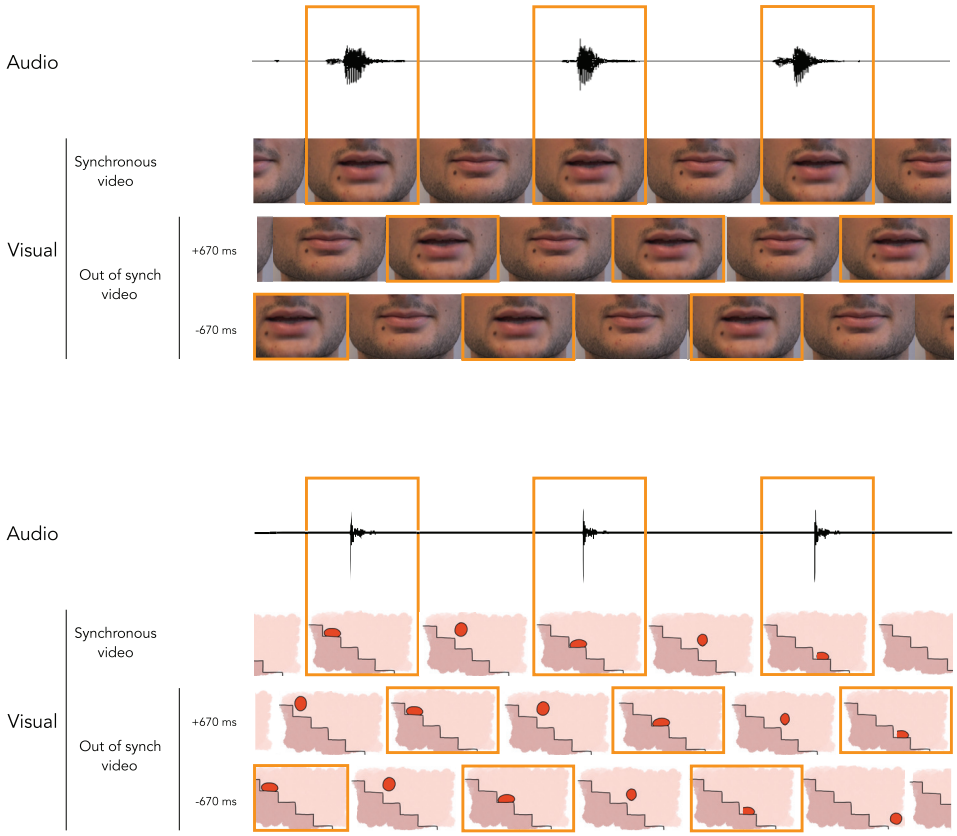


Note. See the online article for the color version of this figure.

Adobe Premiere software. Each stimulus video was split into an audio and a video file. The speech stimulus videos were recorded using a Sony video camera. They were equalized and edited using Audacity software. For each stimulus an out-of-synch video was created by delaying the audio file by 670 ms relative to the onset of the video (for one half of the videos) or advancing it by 670 ms relative to the onset of the video (for the other half); see Figure 2. Due to the introduction of a 670-ms lag, and to avoid cutting the delayed video before the last occurrence of a mouth movement, some videos exceeded the duration of 5 s by a time interval ranging between 172 ms and 840 ms. In the analyses, only the first 5.172 s of all stimulus phases were included.

The two conditions took place on different days, with the condition order being counterbalanced across participants. Each child was assigned to a version consistently across the two conditions: synchronous, in which the reinforcement always appeared on the side of the in-synch video, or asynchronous, in which the reinforcement always appeared on the side of the out-of-synch video.

Figure 2
Video Stimuli: Speech (Top) and Nonsocial (Bottom) Conditions



Note. See the online article for the color version of this figure.

To sum up, we tested: two conditions (speech vs. nonsocial) within participants and two versions (synchronous and asynchronous) between participants. (The same stimuli were used across synchronous or asynchronous version of each condition: In each condition in half the out-of-synch videos the video led the audio and, in the other half, the audio led the video. Preliminary analyses revealed no effect of the direction of the out-of-synch stimulus lag on gaze distribution between reinforced and nonreinforced stimuli, and no interaction between lag direction, group, and area of interest (AOI). For this reason, we do not discuss this aspect of our paradigm any further.)

During each session, the child was seated at a distance of ±60 cm in front of a 16.5-in. monitor (resolution: 1920 × 1080 pixels). Before starting the task, a 5-point calibration procedure designed by Tobii Studio was used. To avoid any bias due to miscomprehension of verbal instructions, especially in nonverbal autistic children, we used an implicit learning paradigm and simply asked children to watch the videos on the screen. The order of the 30 trials was randomized across participants. The audio stimuli were presented at a comfortable SPL (65dB ± 5dB), adapted to each child's sensitivity to noise.

The task was designed using Tobii StudioTM 3.2.1 software and was presented on a computer screen equipped with a Tobii pro X2-60 (Hz) eye-tracker device (Tobii Technology, Inc. Stockholm, Sweden). Areas of interest (AOIs) were also defined using the Tobii Studio 3.2.1 software.

Participants

A fully reliable ASD diagnosis is rarely available before the age of 3 (Lord et al., 2012). Also, at 3 years potential language delays are already manifest and are likely to span until at least the age of 5. As our study targets integration mechanisms whose atypicality is hypothesized to persist since infancy in autistic children, and to compromise their acquisition of speech, 3- to 5-year-old autistic children with minimal or absent expressive language constitute a suitable experimental group to test this hypothesis. The comparison group is composed by TD children matched by chronological age, who have age appropriate speech and language abilities and in whom audio-visual integration processes should be fully operational.

A total of 90 3- to 5-year-old children took part in this study; 41 autistic children (13 girls, 28 boys) and 49 TD children (29 girls, 20 boys). Two autistic children and one TD child were not included in the final data set (the first two were reluctant to watch videos and the latter had a perforated eardrum). The final sample consisted of 39 autistic children (27 boys, 12 girls, age = 55.64 ± 8.63 months; range: 35–72) and 48 TD children (20 boys, 28 girls, age = 42.29 ± 11.10 months; range: 36–72). Children in the ASD group were recruited from the ACTE register of volunteers, through the Center de Référence Autisme "Autrement" and from three functional rehabilitation centers for autistic children. TD children were recruited

from the ACTE register of volunteers, a preschool, and through announcements on the Internet. For all participants, the primary language used at home was French. Informed parental consent and the assent of the children were obtained. The experimental procedure was in accordance with the Declaration of Helsinki, and was approved by the ethics committee of Eramse Hospital. All autistic children had an independent clinical diagnosis of autism made by a multidisciplinary team, in an Autism Reference Center officially entitled to establish a diagnosis of autism. For 35 of these autistic children, the diagnosis was confirmed through the Autism Diagnostic Observation Schedule (Lord et al., 2012; $N = 8$), the Autism Diagnostic Interview-Revised (Rutter et al., 2003; $N = 7$) criteria or both ($N = 20$). The four remaining autistic children also had an independent clinical diagnosis of autism made by an Autism Reference Center, but with tools other than ADI-R or ADOS. For these children the clinical diagnosis for autism was confirmed by a research-accredited ADOS assessor in our team using the ADOS. No child in the TD group had a known history of a neurological or psychiatric condition. Furthermore, the absence of clinically significant levels of autistic symptomatology was confirmed, for all TD children, by the administration, in full, of the ADI-R by the last author (an accredited ADI-R assessor).

As can be seen from Table 1, which summarizes participant characteristics, while our TD and ASD groups had a comparable age range, autistic children were slightly older; this was also the case in Bebko et al. (2006) and Righi et al. (2018). Given that the expected group differences go in the direction opposite to that of this slight age imbalance, it should not affect our results. Assessing nonverbal IQ in ASD, especially in young, minimally verbal children, is fraught with notorious difficulties and biases, and fully reliable measures are difficult to come by (Bishop et al., 2015; Courchesne et al., 2019; Tager-Flusberg et al., 2017). Nonverbal IQ was assessed using the Leiter International Performance Scale–3 (Roid et al., 2013). In spite of older age, scores on this test were lower in the ASD group. The administration of the Leiter scale relies on nonverbal instructions; for this reason, the Leiter is the optimal standardized tool to assess the IQ in nonverbal autistic children. However, the administration had to be stopped in a substantial number of children in our ASD group.

In each condition, children were excluded from analyses because of technical problems ($n = 4$), experimental error ($n = 8$), or withdrawal from one session ($n = 3$), leaving a final sample of 75 children in the speech condition (ASD, $n = 31$; TD, $n = 44$), and 83 in the nonsocial condition (ASD, $n = 36$; TD, $n = 47$). Recall that during the stimulus phase two competing videos—an in-synch and an out-of-synch one—are simultaneously displayed at the screen, that the transition phase corresponds to a blank screen and that only one video is displayed during the reinforced

Table 1
Participant Characteristics

Measure	ASD ($n = 39$)	TD ($n = 48$)	p-value
Age (in months)			
Mean *(SD)*	55.64 (8.63)	42.29 (11.10)	$p = .004$
Age range	35–72	36–72	
Gender			
Female	12	28	
Male	27	20	
Nonverbal IQ[a]			
Mean (*SD*)	79.16 (18.69)	99.30 (7.93)	$p < .001$
Range	41–111	75–117	
First quartile	67	96	
Second quartile	81	100	
Third quartile	96	104	
Mother education[b]			
Primary school	5	0	
Secondary school	10	2	
BA or equivalent	12	14	
MA or equivalent	7	18	
Postgraduate	0	11	
Father education[c]			
Primary school	2	1	
Secondary school	18	6	
BA or equivalent	2	8	
MA or equivalent	4	20	
Postgraduate	5	9	
First-order relative[c]			
With ASD	8	0	
With another neurodevelopmental disorder	0	2	
Second language spoken at home[d]	20	23	

Note. ASD = autism spectrum disorder; TD = typically developing.
[a] Two autistic children refused to participate in the task altogether. There is one missing value in the TD group. [b] Missing data for five children in the ASD group and three in the TD group. There are three children with unknown paternity in the ASD group and one in the TD group. [c] Missing data for two children in the ASD group and three in the TD group. [d] Missing data for five children in the ASD group and two in the TD group.

phase. We reasoned that participants needed to spend at least 50% of the stimulus phase looking at the screen to be able to detect the in- versus out-of-synch temporal alignment. First, we excluded all participants for whom we had gaze recording for at least 50% of the stimulus phase in less than a third of trials: eight participants with ASD in the speech condition, and 10 participants with ASD, and two TD participants in the nonsocial condition. Next, we excluded all trials in which we had gaze recordings from less than 50% of the duration of stimuli phase. In the speech condition this resulted in the exclusion of 33% of trials in ASD group and 10% in the TD group; in the nonsocial condition in the exclusion of also 33% of trials in the ASD group and 9% of trials in the TD group. (Even though our 50% exclusion threshold is theoretically motivated, it is worth noting that all the results reported below remain virtually identical when a less strict threshold, of 25% of fixation is used.)

Data Preparation and Statistical Analyses

Two AOIs were designed and kept constant across the stimuli and the transition phases: *reinforced*, corresponding to the exact zone where the reinforced stimuli was displayed (and hence where the reinforcement appeared in the reinforced phase) and *nonreinforced*, corresponding to the exact zone where the nonreinforced stimulus was displayed. Together, these two AOIs corresponded to 8.86% of the total area of the screen. A third AOI, *screen*, corresponded to the rest of the screen and was used for the analysis of the transition phase. Every 16 ms and for each AOI, we extracted eye-tracking fixation data indicating whether this AOI was active or not (viz., whether a fixation was recorded on that AOI or not); then we averaged fixation values over 100-ms time intervals.

We built three categorical variables: group (ASD vs. TD), AOI (reinforced vs. nonreinforced vs. screen), and version (synchronous vs. asynchronous). We also built a continuous time variable (binned every 100 ms). One of our methodological expectations was that the transition phase, but not necessarily stimulus phase, may be informative as to the children multimodal integration. Accordingly, for each condition, the stimulus, the transition and the reinforcement phases were analyzed separately.

In the stimulus phase, fixations on the screen AOI, rather than on the reinforced or nonreinforced AOI are difficult to interpret, as they can correspond either to gaze transition between the two videos (viz., the reinforced and the nonreinforced AOIs) or to loss of interest from the child. For this reason, in the analysis of the stimuli phases, in both conditions, we used a binomial AOI factor (reinforced vs. nonreinforced). Previous preferential gaze paradigms presuppose that in-synch stimuli exert a familiarity effect in TD children. The prediction, then, is that TD children should be attracted by in-synch stimuli—which correspond to the reinforced AOI in our synchronous condition and to the nonreinforced AOI in our asynchronous condition—while no such preference should emerge in autistic children. Note that, in our experimental design, the addition of a reinforcement procedure in the transition phase may progressively interfere with initial preference for in-synch stimuli in the stimulus phase. However, independently of the reinforcement and as argued in the Introduction, it is also possible that in TD and ASD children alike, gaze trajectories are mainly driven by salient changes in the stimuli, irrespective of the temporal alignment of the audio and video components. Analysis of fixation curves between reinforced and nonreinforced AOIs should allow to clearly visualize whether salient changes determine gaze movements.

In the transition phase, preferential fixations on the reinforced AOI correspond to the correct anticipation of the reinforcement based on the temporal alignment of the stimuli; in this case, therefore, it is important to compare fixations on the reinforced AOI with those on the rest of the screen, namely on the screen. It is also informative to include the nonreinforced AOI in the analyses, as fixations on this AOI may reflect erroneous anticipation of the reinforcement.

Finally, if TD children and autistic children find the reinforcement videos attractive to the same extent, during the reinforcement presentation phase no difference in fixations on the reinforcement AOI (which corresponds to the reinforcement animation video) should emerge between groups.

All statistical analyses were implemented in R (R Core Team, 2016). Average fixation values per 100 ms were analyzed implementing linear generalized multilevel regressions using the lme4 package (Bates et al., 2015). All multilevel regressions included by-participant random intercepts, and, when possible, time by-participant random slopes (this was the most complex random structure to allow model convergence). Relative to traditional analyses of variance, used in the previous studies on multimodal integration discussed above (Bebko et al., 2006; Righi et al., 2018), such multilevel models have the advantage of minimizing the risk of Type I error (e.g., Barr et al., 2013). Significance of the fixed effects was assessed by performing stepwise likelihood ratio tests in which a model containing the fixed effect is compared with another model without it but with an otherwise identical random effect structure. We used the lsmeans package (Lenth, 2016) for posthoc comparisons of least square-means (lsmeans) and estimations of slopes (lstrends), with Tukey adjustment for multiple comparisons. In order to capture nonlinear time curves, we also used generalized additive multilevel models, using the mgcv package (Wood, 2017). Raw data are available as online supplementary materials.

Results

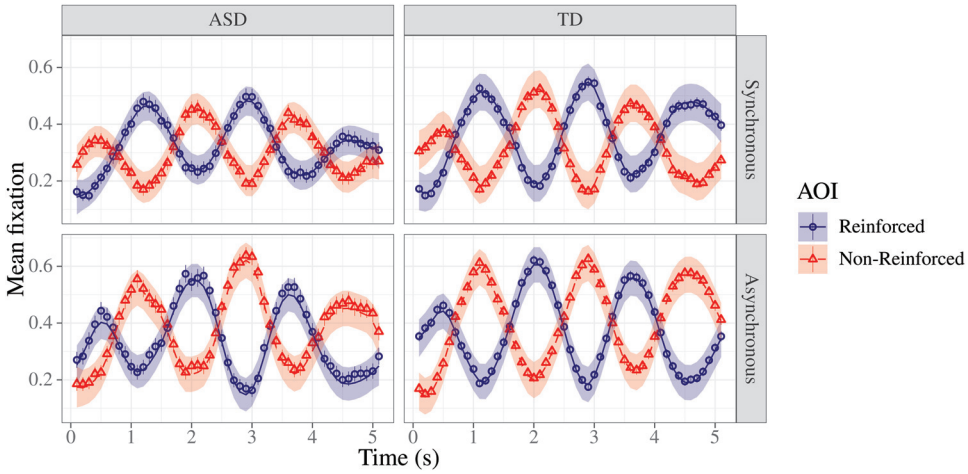
Speech Condition

Stimulus Presentation Phase

Stepwise comparisons of multilevel linear regressions (with time by-participant random slopes) revealed an effect of AOI, version and group, as well as the associated interactions (all $p < .001$). Next, we conducted posthoc pairwise comparisons of the triple AOI $\times$ Group $\times$ Version interaction. First, there were no group differences, in either version, in the amount of fixations on the reinforced or on the nonreinforced AOI (all $p = 1$). Second, the reinforced AOI attracted more fixations than the nonreinforced AOI in the synchronous version in both groups (ASD: $\beta = 0.02$; $se = 0.05e^{-1}$; $p = .003$; TD: $\beta = 0.04$; $se = 0.04e^{-1}$; $p < .001$), while in the asynchronous version, the reverse pattern was true (ASD: $\beta = -0.04$; $se = 0.06e^{-1}$; $p < .001$; TD: $\beta = -0.04$; $se = 0.03e^{-1}$; $p < .001$). Recall that the reinforced AOI corresponded to the in-synch video in the synchronous version, but to the out-of-synch video in the asynchronous version. So, all other things being equal, these results indicate a preference for the in-synch video in both groups.

Figure 3

Speech Condition, Stimulus Presentation Phase



Note. ASD = autism spectrum disorder; TD = typically developing; AOI = area of interest. Mean fixation values (per 100 ms bins) per AOI and fitted curves; vertical bars represent standard errors of means and shadow ribbons represent 95% confidence intervals. See the online article for the color version of this figure.

Recall, however, that each stimulus in the speech conditions contained three salient mouth closures (see Figure 2). It is therefore plausible that these three salient articulatory movements were a driving factor in the distribution of visual fixations in both groups. To better understand visual fixation trajectories we fitted, for each version and group, a generalized multilevel linear additive model, with an AOI parametric term, time, and time by AOI fixed smooth terms, and time by-participant and by-item random smooths (see Table B1 in Appendix B). Fitted curves are plotted in Figure 3, together with mean fixations per time point; these results provide clear indication that fixation distribution during the presentation of speech stimuli was highly periodic—with three phases that correspond to the three salient articulatory movements in the stimulus videos—irrespective of the version and group.

Transition Phase

Stepwise comparisons of multilevel linear regressions (with time by-participant random slopes) on fixation during transition presentation in the speech condition revealed an effect of AOI, version, group, and time, as well as the associated interactions (all $p < .001$). Fitted slopes are displayed in Figure 4A; in both ASD and TD groups, and in both versions, fixations on either the reinforced or nonreinforced AOI drastically drop around the half of the 1-s transition period, while those on the rest of the screen AOI increase. Posthoc comparisons indicated that, in both versions and groups, the negative slopes of reinforced and nonreinforced AOIs are significantly different from the positive slopes of fixations of the rest of the screen (all $p < .001$). This X-shaped fixation patterns between, on the one hand, the reinforced and nonreinforced AOI and, on the other hand, the rest of the screen make it somehow difficult to assess potential differences between categorical predictors and their interactions. For this reason, we censored all the data for the transition period below .5 s, and implemented a linear multilevel model with AOI × Version × Group fixed terms and by-participant random intercepts. The effects of all categorical predictors are displayed in Figure 4B. Posthoc comparisons indicated that, in both versions, TD children fixated all AOIs more than autistic children (all $p \leq .001$). In TD children, there was a clear preference for the reinforced AOI over the nonreinforced AOI (synchronous: $\beta = 0.4; se = 0.01; p < .001$; asynchronous: $\beta = 0.04; se = 0.01; p < .001$), as well as over the rest of the screen (both $p < .001$). In the ASD group, there was a similar preference for the reinforced over the nonreinforced AOI in the asynchronous version ($\beta = 0.06; se = 0.02; p = .001$), but not in the synchronous version ($p = .54$). However, in both versions, autistic children fixated more the reinforced AOI than the rest of the screen (both $p < .001$).

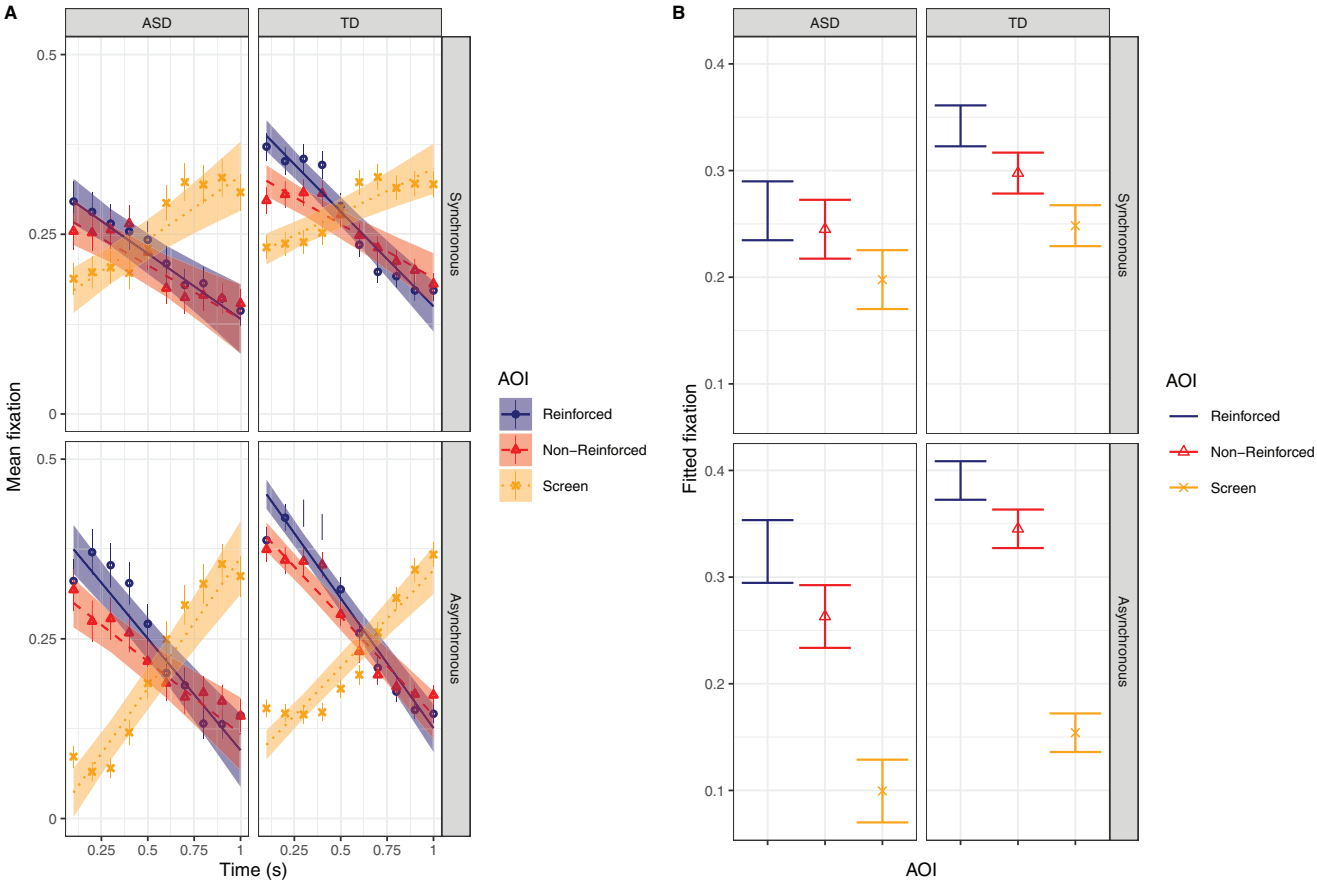
In sum, during the first half of the transition period, both autistic children and TD assigned to the version in which the out-of-synch was reinforced, preferentially gazed toward the part of the screen on which the reinforcement video would appear. In the version in which it was the in-synch video that was reinforced, only TD children preferentially gazed toward the part of the transition screen on which the reinforcement video would appear.

Nonsocial Condition

Stimuli Phase

Stepwise comparisons of multilevel linear regressions (with by-participant random intercepts) on fixation during stimuli presentation in the nonsocial condition revealed an effect of AOI, version, and group, as well as of the associated interactions (all $p < .001$). Posthoc pairwise comparisons revealed, in TD children, a lower amount of fixations on the reinforced than on the nonreinforced AOI in the synchronous version ($\beta = -0.02; se = 0.03e^{-1}; p < .001$), and a higher amount of fixations on the reinforced than on the nonreinforced AOI in the asynchronous version ($\beta = 0.03; se = 0.03e^{-1}; p < .001$). Furthermore, in the asynchronous version autistic children fixated less the reinforced AOI

Figure 4
Speech Condition, Transition Phase



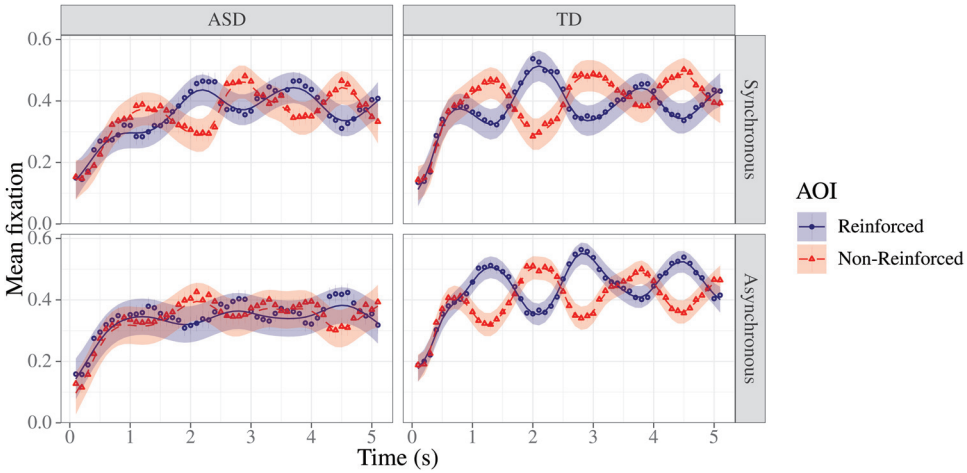
Note. A. Mean fixation values (per 100 ms bins) and fitted fixation slopes. Vertical bars represent standard errors of means and shadow ribbons represent 95% confidence intervals. B. Effects of the categorical predictors (first 0.5 s); error bars represent 95% confidence intervals. ASD = autism spectrum disorder; TD = typically developing; AOI = area of interest. See the online article for the color version of this figure.

than TD children ($\beta = -0.1; se = 0.03; p = .016$). No other contrast was significant (all $p > .92$). While these differences may be somewhat difficult to interpret, recall that nonsocial stimuli also have a periodic dimension (see Figure 2). We fitted, for each version and group, a generalized multilevel linear additive model, with an AOI parametric term, time, and time by AOI fixed smooth terms, and time by-participant, and by-item random smooths (see Table B2 in Appendix B). The fitted curves are plotted in Figure 5, together with mean fixations values. These data indicate that TD children's fixation distribution was clearly determined by periodic events in the video stimuli. Interestingly, patterns of fixation were much less systematic in autistic children, with a much less clearer difference between the two AOIs.

Transition Phase

Stepwise comparisons of multilevel linear regressions (with time by-participant random slopes) on fixation during the transition phase in the nonsocial condition revealed an effect of AOI, version, group, and time, as well as the associated interactions (all $p < .001$). Fitted slopes are displayed in Figure 6A. As in the speech condition, in both ASD and TD groups, and in both versions, fixations on either the reinforced or nonreinforced AOI drop around the half of the 1 s transition period, while those on the rest of the screen AOI increase. Figure 6A also indicates an overall preference for the reinforced AOI. To better visualize the effects of categorical predictions, we censored all the data for the transition period below .5 s, and implemented a linear multilevel model with AOI × Version × Group fixed terms and by-participant random intercepts. The effects of all categorical predictors are displayed in Figure 4B. Posthoc comparisons indicated that TD children fixated more the reinforced and the nonreinforced AOIs than autistic children (all $p < .001$); there was no difference in fixations on the rest of the screen (both $p > .32$). In the synchronous version in both groups there were more fixations on the reinforced than on the nonreinforced AOI, (ASD: $\beta = 0.07$; $se = 0.01e^{-1}$; $p < .001$; TD: $\beta = 0.13$; $se = 0.01e^{-1}$; $p < .001$). On the contrary, in the asynchronous version, in both groups, the reinforced AOI was less fixated than the nonreinforced AOI (ASD: $\beta = -0.08$; $se = 0.2$; $p < .001$; TD: $\beta = -0.04$; $se = 0.01$; $p < .001$). In all the versions and groups, the reinforced AOI was more fixated than the screen AOI (all $p < .001$). In sum, in the nonsocial condition, both autistic children and TD preferentially gazed toward the part of the screen on which in-synch stimulus was displayed, whether it was

Figure 5
Nonsocial Condition, Stimulus Presentation Phase



Note. Mean fixation values (per 100 ms bins) per AOI and fitted curves; vertical bars represent standard errors of means and shadow ribbons represent 95% confidence intervals. ASD = autism spectrum disorder; TD = typically developing; AOI = area of interest. See the online article for the color version of this figure.

the side that was subsequently reinforced (in the synchronous version) or not (in the asynchronous version).

Attention to the Reinforcement Animation

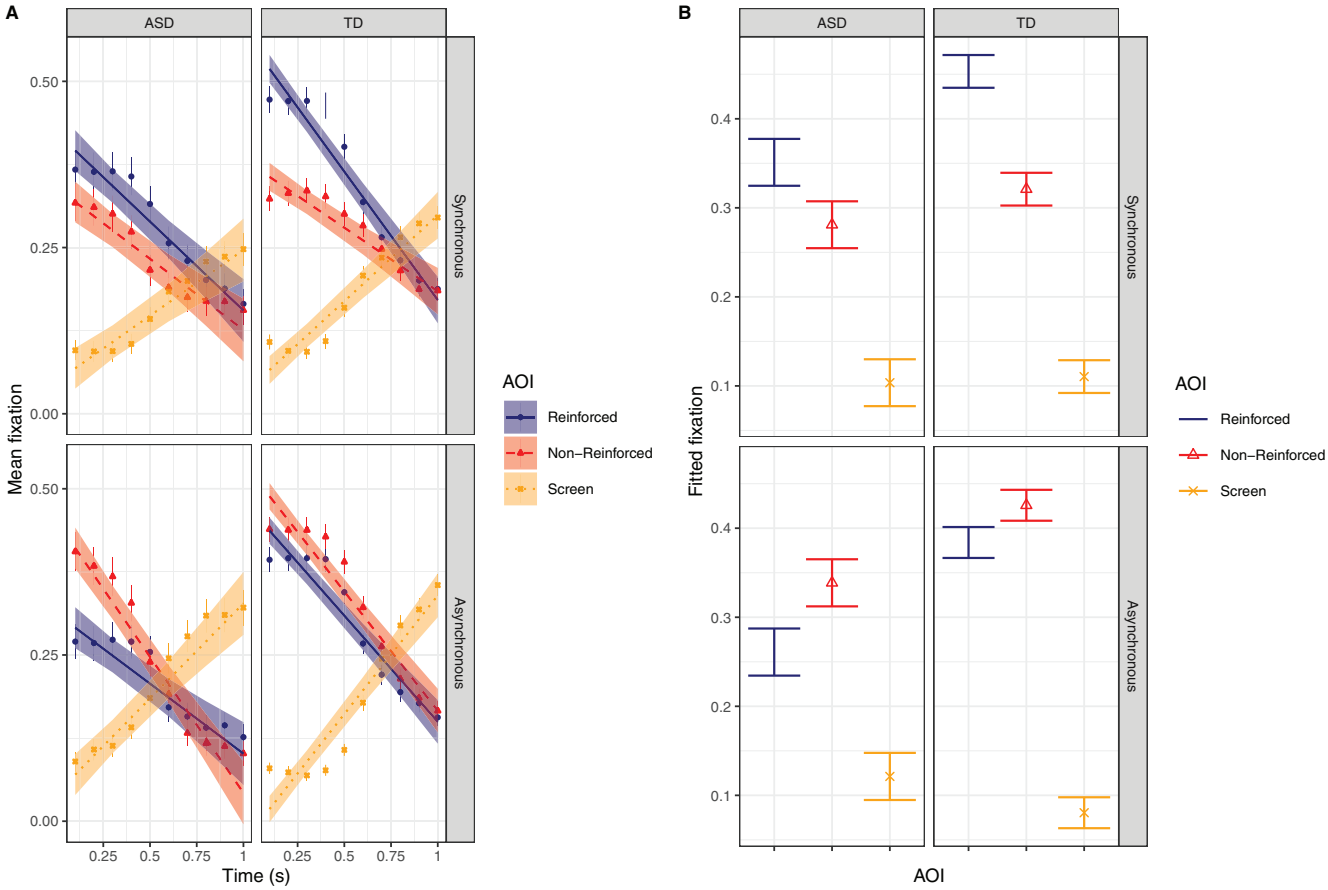
In order to assess whether children in the two groups were visually attracted by the reinforcement animations to the same extent, we analyzed fixations on the reinforcement animation fitting linear multilevel models with by-participant and by-item random intercepts. The addition of the group factor did not improve the model fit in the speech condition (p = .37) or in the nonsocial condition (p = .21). That is, there are no grounds for assuming that autistic children were less attracted by the reinforcement animations than their TD peers.

Discussion

The main objective of this article was to implement a novel methodology to investigate potential difficulties young, minimally or nonverbal autistic children could experience in mapping articulatory mouth movements on the corresponding acoustic signal. There is converging evidence that audio-visual integration in speech may be atypical in autism (Bahrick & Todd, 2012; Feldman et al., 2018; Stevenson et al., 2014). However, most of the existing literature on multimodal integration in autism focuses on highly verbal adults, and tells little about the processing profiles at an earlier developmental stage, in children in whom language delays become evident. Yet, as emphasized by Bahrick and Todd (2012), it is crucial to gather reliable evidence about multimodal integration in autism at the earliest developmental stage possible. Two studies (see Bebko et al., 2006; Righi et al., 2018) took up the challenge and included younger autistic children, using the preferential gaze paradigm to investigate multimodal integration. These studies provide important indication that audio-visual integration in speech may be compromised in ASD, but they also failed to uncover clear-cut group differences between autistic children and their TD peers. This absence of group effects may be partly due to the fact that in both studies a significantly asymmetrical fixation distribution between in- and out-of-synch stimuli was the only way to detect successful audio-visual integration.

Albeit preferential gaze methodology proved reliable in infants (e.g., Hillairet de Boisferon et al., 2017; Kuhl & Meltzoff, 1984; Patterson & Werker, 2003), the results of our speech condition indicate that it is not optimal to investigate audio-visual integration in older children, autistic or not. The first conclusion to emerge from our analyses of visual exploration of in- and out-of-synch video stimuli is that a significant preference for familiar in-synch or for novel out-of-synch videos should not be presupposed in investigating audio-visual integration—in autism and in general. In our speech condition, the distribution of fixations in the stimuli phase shows a preference for the in-synch video in both groups (see, Table 2), but this familiarity effect is not the main factor determining children's gaze trajectories. A fine-grained analysis of eye fixation trajectories over time reveals that even with simple mouth stimuli, stripped of potentially distracting factors such as eyes or other facial features, children's gaze is strongly influenced by the periodic and salient mouth movements; see Figure 3. It is worth emphasizing that our data show more than a back-and-forth alternance between the two videos. The consistently periodic gaze pattern is clearly determined by the periodic features of the stimuli. Importantly, the attraction periodic articulatory movements exert on children's gaze does not depend on the in- or out-of-synch audio-visual alignment of the video. This strong attraction by the periodic aspects of the videos may obfuscate any potential group differences—in our data, but also in other studies, where it was not analyzed. These results should prompt researchers not to rely on preferential gaze paradigms without a clear reflection on the aspects of the stimuli that may attract the participants' gaze. They also underline the advantages of using time series in order to unveil gaze trajectories that are concealed in more traditional

Figure 6
Nonsocial Condition, Transition Phase



Note. A. Mean fixation values (per 100 ms bins) and fitted fixation slopes. Vertical bars represent standard errors of means and shadow ribbons represent 95% confidence intervals. B. Effects of the categorical predictors (first 0.5 s); error bars represent 95% confidence intervals. ASD = autism spectrum disorder; TD = typically developing; AOI = area of interest. See the online article for the color version of this figure.

analyses, which collapse proportions of fixations during a given time window (compare with Bebko et al., 2006; Righi et al., 2018).

We found no group difference in the fixation patterns on speech stimuli, as children in both ASD and TD group exhibited a highly periodic gaze distribution, with a preference for in-synch stimuli. However, the fact that in our experimental design preferential gaze was reinforced provides new insights into audio-visual processing in young, nonverbal autistic children. Fixation patterns collected during the transition phase of the experimental trials provide a fuller view of the child's capacity to distinguish between in-synch and out-of-synch video recordings than traditional paradigms, which are limited to comparing fixations on in- and out-of-synch videos. As summarized in Table 2, results of fixation trajectories in the transition phase of the speech condition showed that TD children were capable of anticipating the apparition of a reinforcement based on the temporal (mis-)alignment of mouth movements with the corresponding speech signal. This was the case in the synchronous version, in which the reinforcement appeared on the side of the in-synch video, and in the asynchronous version, in which the reinforcement appeared on the side of the out-of-synch video. Recall that the reinforcement primed either the aligned or the misaligned video—depending on the version to which the participant was assigned. Therefore, the fact that in both versions the reinforced AOI was more likely to be fixated is most probably due to anticipation of the reinforcement video. In other words, TD children are capable of learning the association between a video reinforcement and the temporal alignment of the audio and the video components of the speech stimuli.

As for autistic children, they also preferentially gazed toward the part of the screen on which the reinforced video would appear in the asynchronous version, but not in the synchronous version; see Table 2. Furthermore, whereas autistic children visually explored the stimulus videos and the reinforcement animations to the same extent as TD children, they also spent, overall, less time fixating the different AOIs during the transition phase. A plausible explanation for such visual disengagement from the screen during the transition phase in the ASD group is that autistic children have more difficulty in anticipating the apparition of the reinforcement video based on the audio-visual alignment of the stimuli. But while the absence of difference between reinforced and

Table 2
Stimulus and Transition Phases: Differences in Fixation Between Reinforced and Nonreinforced AOIs per Condition, Version and Group

Version	Stimulus phase		Transition phase	
	ASD	TD	ASD	TD
	Speech condition			
Synchronous (In-sync reinforced)	Reinforced > Nonreinforced	Reinforced > Nonreinforced	Reinforced = Nonreinforced	Reinforced > Nonreinforced
Asynchronous (Out-of-sync reinforced)	Reinforced < Nonreinforced	Reinforced < Nonreinforced	Reinforced > Nonreinforced	Reinforced > Nonreinforced
	Nonsocial condition			
Synchronous (In-sync reinforced)	Reinforced = Nonreinforced	Reinforced < Nonreinforced	Reinforced > Nonreinforced	Reinforced > Nonreinforced
Asynchronous (Out-of-sync reinforced)	Reinforced = Nonreinforced	Reinforced > Nonreinforced	Reinforced < Nonreinforced	Reinforced < Nonreinforced

Note. ASD = autism spectrum disorder; TD = typically developing; AOI = area of interest.

nonreinforced AOI in the synchronous version is also consistent with the idea of a lower ability to anticipate the location of the reinforcement based on the audio-visual temporal (mis)alignment, it is less clear why an anticipatory preference for reinforced AOI did emerge in the asynchronous version. A plausible interpretation for this difference between versions is that in the ASD group the preference for the reinforced AOI in the asynchronous version is due to those children who already developed some expressive language. Figures A1-A4, in Appendix A, display analyses of the eye-tracking data when those autistic children who had some expressive language at the time of testing are excluded. As shown in Figure A2, subtracting those autistic children who have some functional language from the analyses annihilates the difference in fixations between the reinforced and the nonreinforced AOIs in the asynchronous condition. These data suggest that young autistic children who do not have functional language experience difficulties in anticipating the location of the reinforcement based on the audio-visual temporal (mis)alignment.

The differences in visual exploration between autistic and TD children that we found in the transition phase of the speech condition constitute new evidence, probably less equivocal than what has been gathered in the literature so far, that audio-visual integration in speech is not entirely operational in young, nonverbal or minimally verbal autistic children. Our data thus indicates that autistic children do not manage to link (or do so to a much lesser extent than their TD peers) the temporal alignment of the audio and visual components of the speech signal with the location of the reinforcement video. To be sure, lower anticipation of the reinforcement video could be due to a lower capacity to implicitly learn to use temporal alignment to anticipate subsequent events. However, current evidence rather robustly indicates that implicit, associative learning mechanisms are intact in autism (e.g., Brown et al., 2010; Haebig et al., 2017; see the meta-analyses in Foti et al., 2015; Obeid et al., 2016). It is likely, then, that the group differences we uncovered are due to difficulties autistic children have in temporally binding mouth movements with the acoustic signal —and thus in distinguishing between in- and out-of-synch stimuli.

Our second objective was to apply the method we used to investigate audio-visual integration in speech to structurally similar nonsocial stimuli. Let us begin, again, by discussing the results observed in our TD group. In the stimulus phase of the nonsocial condition, TD children displayed more fixations on the nonreinforced than on the reinforced AOI in the synchronous version, but more fixations on the reinforced AOI than on the nonreinforced AOI in the asynchronous version; see Table 2. This fixation pattern could be taken to indicate a preference for the out-of-synch video in both versions. Under this interpretation, in TD children preferential fixations would be driven by the novelty of out-of-synch nonsocial stimuli, but by the familiarity of in-synch speech stimuli. However, such a dissociation, based on the nature of the stimuli, would rather be difficult to defend on principled grounds. Importantly, when a time-course analysis of fixation distribution is adopted, a clear periodic gaze pattern also emerges in the nonsocial condition, showing that a back-and-forth alternance between the two videos is determined by the three periodic events in the stimuli; see Figure 5. Therefore, in paralleling the speech condition, the salient changes in the stimulus videos appear to be the most important factor driving TD children's visual attention.

Turning to the transition phase, a somehow unexpected, but interesting contrast emerges between the transition phases in speech and nonsocial conditions. As we just saw, in the transition phase of the speech condition, TD children displayed visual preference for the reinforced AOI in both versions, indicating that they were primed by the reinforcement, irrespective of whether it was the in- or the out-of-synch video that was reinforced. In the transition phase of the nonsocial condition, by contrast, TD children displayed a significant preference for the reinforced AOI in the synchronous version, but for the nonreinforced AOI in the asynchronous version; see, Table 2. This latter gaze distribution pattern suggests that TD children were more prone to erroneous anticipation in the asynchronous version. One source of the difference between the two conditions—and a potential limitation of our material—could be that the stimuli in the nonsocial condition represented meaningful, short stories, while those in the speech conditions did not contain any narrative detail to delve on. Moreover, the nonsocial stimuli contained many visual features that were not informative as to the audio-visual alignment, which could have made the detection of (a)synchrony more difficult. Finally, in

the speech condition, the reinforcement videos consisted in animations that pictured unexpected events on the actors' mouths, while in the nonsocial conditions, the reinforcement depicted is an unexpected culmination of the periodic event represented in the video stimuli. It is therefore possible that in the nonsocial condition, children's attention was less drawn to temporal alignment, which, by contrast, constituted the main feature in the speech condition. Further studies are clearly needed to explore the features of visual stimuli that may influence audio-visual integration in children.

Let us turn to the results in the ASD group, starting this time with findings from the transition phase. The distribution of fixations in autistic children follows the same pattern as TD children with a significant preference for the reinforced AOI in the synchronous version, but the nonreinforced AOI in the asynchronous version; see Table 2. At the first glance, these results seem to indicate that, as their TD peers, autistic children correctly anticipate the reinforced video in the synchronous version, but are more error prone in the asynchronous version. However, two pieces of evidence indicate that, relative to their TD peers, audio-visual integration in autistic children is also disrupted in the nonsocial condition. First, independently of the version, autistic children displayed less fixations on the reinforced and nonreinforced AOIs than TD children in the transition phase of the nonsocial condition. As in the speech condition, lower exploration of these AOIs during the transition phase likely reflects difficulties in anticipating the location of the reinforcement videos based on audio-visual properties of the stimuli. Second, while fixation distribution on the nonsocial stimuli was clearly periodic in TD children, thus driven by the salient periodic events in the video animations, this was much less the case in autistic children; see Figure 5.

Klin et al. (2009) showed that multimodal integration may be easier for autistic children in stimuli in which the co-occurrence of video and audio events is supported by particularly salient physical events and acoustic signals. Given that the videos of speech stimuli were stripped of the upper part of the face, the mouth movements in the speech stimuli offered a clearer locking opportunity with the acoustic signal than the nonsocial stimuli, which, albeit very simple, did contain more detail and had more variability in the periodic event-sound pairs. As explained in the Introduction, we restricted the visual part of our stimuli to the mouth region to avoid potential confounds, such as atypical face exploration or face aversion in autistic children. Of course, we cannot rule out that audio-visual integration could be easier for autistic children who could benefit from redundancy cues to audio-visual alignment from upper parts of the face. It is much more likely, though, that real-life audio-visual integration in speech is more difficult to process for autistic children, precisely because of the speed, the number and the fine nature of articulatory movements involved in a speaking face (see, also Bahrick & Todd, 2012). In sum, our results complement bottom-up models, which hypothesize a sensory processing basis for language and communication deficits in ASD (Robertson & Baron-Cohen, 2017; Stevenson et al., 2014, 2018). Even though autism is indisputably characterized by an atypical processing of socially meaningful information, speech processing in young autistic children may be further impacted by difficulties in integrating audio and visual components of complex, rapidly evolving events.

Our study clearly calls for further, longitudinal studies, which should aim at assessing the link between the emergence of language and multimodal integration in ASD. Albeit most of our autistic children were totally nonverbal (see Appendix A), they were also in the age range during which language abilities may emerge in previously nonverbal autistic children, in an often quite abrupt and not entirely predictable manner (e.g., Anderson et al., 2007; Ellis Weismer & Kover, 2015; Thurm et al., 2015; Wodka et al., 2013). As can be seen from Appendix A, all contrasts and fixations curves remain identical when only fully nonverbal children are kept in the analysis, with the exception of the difference between fixations on the reinforced versus nonreinforced AOI in the asynchronous version of the speech condition (see above). Nevertheless, it is possible that some of these autistic children who displayed no functional language at the moment of testing were actually on the verge of acquiring linguistic skills.

On a more general, methodological note, the reinforcement paradigm and the analytic methods used in this article could prove valuable for any type of experimental paradigm that relies on preferential gaze. There is no entirely principled way to predict, especially in clinical populations, the relative weights of familiarity vs novelty effects or whether such effects may significantly outweigh other factors in driving participants' visual attention. The introduction of a reinforcement allows to assess participants' sensitivity to the difference between the stimuli without presupposing that they should be more attracted by novel or, on the contrary, by familiar properties. A related point, which we already touched upon above, is that some properties of dynamic stimuli (e.g., salient articulatory movements) may drive visual attention independently of the experimentally manipulated dimensions (e.g., audio-visual alignment). In order to carefully delineate all the factors that may attract participants' visual attention, the fixation trajectories during stimuli presentation phases should be more systematically analyzed.

The methodological points just discussed could also be relevant for a promising protocol, which could help assessing multisensory integration in nonverbal autistic children, put forth by Bahrick et al. (2018). This paradigm, which, in addition to audio-visual integration in social and nonsocial domains, also targets visual attention maintenance and speed of disengagement, consists in out- and in-synch stimuli, social and nonsocial, presented on each side of a window that sometimes contain a distractor event. In their result analysis, Bahrick et al. (2018) use (manually coded) proportions of looks. Therefore, this paradigm also relies on an asymmetric fixation distribution to measure the detection of audio-visual synchrony. The introduction of a reinforcement, along, perhaps, with a more precise time-series analysis of gaze patterns, could increase the robustness and the reliability of Bahrick et al.'s (2018) protocol.

A clear limitation of the present study lies in the relatively high number of autistic children or trials within the ASD group we had to exclude because of lack of gaze recordings during the stimuli phase. This reduction of the initial sample of data does prompt some caution as to the generalization of our results. Missing eye-tracking data may be caused by a lack of interest in the experimental procedure or by some aversion induced by features of our stimuli; either way, the relationship with audio-visual integration skills can only be hypothetical. At the same time, because our exclusion criteria were particularly stringent, the amount of visual exploration of the stimuli in all trials kept for analyses ensures that all participants had sufficient opportunity to detect audio-visual (mis)alignment. This is another respect in which our analytical method allows a more reliable

interpretation of group differences than the window analysis used by previous studies of audio-visual integration in young autistic children. To compare, in Righi et al. (2018) the threshold for the exclusion of a trial was set at less than 500 ms of fixation at the screen over 14-s long trials, namely less than 4% of the total trial duration, while Bebko et al. (2006) averaged looking times across trials. Given that all the results submitted to analysis by these authors were aggregated as proportional gaze distribution per trial, there is a risk to conflate lack of preferential gaze to in-synch videos with a reduced opportunity to detect (a)synchrony due to an overall low fixation on the stimuli.

Relatedly, due to difficulty in obtaining reliable nonverbal IQ scores for many of our autistic participants, our experimental groups were not matched on mental age, which may be seen as another limitation of our article. Let us stress, however, that lack of reliable standardized IQ scores is a characteristics inherent in testing nonverbal or minimally verbal autistic children. In most autistic children who have no functional language, the administration of nonverbal IQ tests, even that of the Leiter (Roid et al., 2013) which does not involve verbal instructions, is very difficult and the collected scores are often unreliable. These testing difficulties may also bias estimates of nonverbal IQs (see, e.g., Bishop et al., 2015; Courchesne et al., 2019; Tager-Flusberg et al., 2017). Furthermore, there is robust evidence that nonverbal IQ or the presence of intellectual delay do not correlate well with adaptive function scores in autism, which may remain low even though IQ scores are in the superior range (e.g., Alvares et al., 2020; Pathak et al., 2019). This, of course, makes it rather difficult to find reliable matching criteria when conducting research on young nonverbal autistic children. However, nonverbal or minimally verbal autistic children represent a very important proportion, around 60%–70%, of 3- to 6-year-olds on the spectrum, and the absence of robust matching criteria should not prevent the scientific community from gathering experimental evidence on this group. Children in our autistic group were also slightly older than TD children (as was also the case in Bebko et al., 2006 and Righi et al., 2018). However, as this slight age imbalance goes in the opposite direction to the group effect we found, it would be problematic only if the detection of audio-visual alignment in our eye-tracking tasks would somehow deteriorate with chronological age.

Finally, the slopes in the transition phase across time, in both conditions, are probably indicative that 1 s is an overly long transition period, so that initial anticipatory looks on the area where the reinforced video is expected to appear are followed by saccades to other parts of the white screen. Still, the robustness of initial gaze on the reinforced AOI—and also to the nonreinforced AOI, indicative of an erroneous anticipation—is remarkable, especially given the small proportion of the screen (8.86%) these AOIs represented and the fact that in the transition period they are not visually delimited in any way.

Conclusion

Eye-tracking technology becomes increasingly available and flexible, providing researchers with a precious window on cognitive processing in young children—and especially those in whom the collection of more traditional behavioral data is rendered virtually impossible by the absence of language or a low developmental level. The study of audio-visual integration in young autistic children, presented above, is a perfect case in point. Using an entirely nonverbal paradigm, our eye-tracking study confirmed that autistic children experience difficulties in temporally binding audio and visual components of video stimuli. As mapping the speech signal on articulatory mouth movements is a crucial milestone in typical language acquisition, a lower ability to integrate multimodal information may contribute to language onset delays that are frequently attested in autism. Importantly, our article also indicates that the difficulty autistic children appear to have in matching audio and visual signals arises in speech and nonspeech stimuli alike. This finding fully warrants further exploration of bottom-up, sensory based models, which link higher-order linguistic and cognitive deficits in ASD to an atypical sensory processing.

Our results also show that the growing enthusiasm toward eye-tracking methods should be somewhat tempered with a careful reflection on the interpretation of the elicited gaze patterns. Simple preferential gaze paradigms may not be ideal in that respect, at least not with children older than 2 years. Methods such as the reinforcement paradigm implemented in the study reported in this article may reduce interpretation biases and optimize the collection of gaze data.

References

Alvares, G. A., Bebbington, K., Cleary, D., Evans, K., Glasson, E. J., Maybery, M. T., Pillar, S., Uljarević, M., Varcin, K., Wray, J., & Whitehouse, A. J. O. (2020). The misnomer of 'high functioning autism': Intelligence is an imprecise predictor of functional abilities at diagnosis. *Autism*, *24*(1), 221–232. https://doi.org/10.1177/1362361319852831

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*.

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., Welch, K., & Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, *75*(4), 594–604. https://doi.org/10.1037/0022-006X.75.4.594

Autism and Developmental Disabilities Monitoring Network. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years – Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, *65*(3), 1–23.

Baghdadli, A., Michelon, C., Pernon, E., Picot, M.-C., Miot, S., Sonié, S., Rattaz, C., & Mottron, L. (2018). Adaptive trajectories and early risk factors in the autism spectrum: A 15-year prospective study. *Autism Research*, *11*(11), 1455–1467. https://doi.org/10.1002/aur.2022

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*(3), 99–102. https://doi.org/10.1111/j.0963-7214.2004.00283.x

Bahrick, L. E., & Todd, J. T. (2012). Multisensory processing in autism spectrum disorders: Intersensory processing disturbance as a basis for atypical development. In B. F. Stein (Ed.), *The new handbook of multisensory processes* (pp. 657–674). MIT Press.

Bahrick, L. E., Todd, J. T., & Soska, K. C. (2018). The Multisensory Attention Assessment Protocol (MAAP): Characterizing individual differences in multisensory attention skills in infants and children and relations with language and cognition. *Developmental Psychology*, *54*(12), 2207–2225. https://doi.org/10.1037/dev0000594

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry*, *47*(1), 88–98. https://doi.org/10.1111/j.1469-7610.2005.01443.x

Bennett, T. A., Szatmari, P., Georgiades, K., Hanna, S., Janus, M., Georgiades, S., Duku, E., Bryson, S., Fombonne, E., Smith, I. M., Mirenda, P., Volden, J., Waddell, C., Roberts, W., Vaillancourt, T., Zwaigenbaum, L., Elsabbagh, M., & Thompson, A. (2015). Do reciprocal associations exist between social and language pathways in preschoolers with autism spectrum disorders? *Journal of Child Psychology and Psychiatry*, *56*(8), 874–883. https://doi.org/10.1111/jcpp.12356

Bishop, S. L., Farmer, C., & Thurm, A. (2015). Measurement of nonverbal IQ in autism spectrum disorder: Scores in young adulthood compared to early childhood. *Journal of Autism and Developmental Disorders*, *45*(4), 966–974. https://doi.org/10.1007/s10803-014-2250-3

Brown, J., Aczel, B., Jiménez, L., Kaufman, S. B., & Grant, K. P. (2010). Intact implicit learning in autism spectrum conditions. *The Quarterly Journal of Experimental Psychology*, *63*(9), 1789–1812. https://doi.org/10.1080/17470210903536910

Campbell, D. J., Shic, F., Macari, S., & Chawarska, K. (2014). Gaze response to dyadic bids at 2 years related to outcomes at 3 years in autism spectrum disorders: A subtyping analysis. *Journal of Autism and Developmental Disorders*, *44*(2), 431–442. https://doi.org/10.1007/s10803-013-1885-9

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLOS Computational Biology*, *5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Courchesne, V., Girard, D., Jacques, C., & Soulières, I. (2019). Assessing intelligence at autism diagnosis: Mission impossible? Testability and cognitive profile of autistic preschoolers. *Journal of Autism and Developmental Disorders*, *49*(3), 845–856. https://doi.org/10.1007/s10803-018-3786-4

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*(1), 69–86. https://doi.org/10.1080/09541449108406220

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective pragmatic impairment in autism spectrum disorder: Indirect requests versus irony. *Journal of Autism and Developmental Disorders*, *48*(9), 2938–2952. https://doi.org/10.1007/s10803-018-3561-6

Ellis Weismer, S., & Kover, S. T. (2015). Preschool language variation, growth, and predictors in children on the autism spectrum. *Journal of Child Psychology and Psychiatry*, *56*(12), 1327–1337. https://doi.org/10.1111/jcpp.12406

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *95*, 220 –234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*, *203*(2), 381–389. https://doi.org/10.1007/s00221-010-2240-4

Foti, F., De Crescenzo, F., Vivanti, G., Menghini, D., & Vicari, S. (2015). Implicit learning in individuals with autism spectrum disorders: A meta-analysis. *Psychological Medicine*, *45*(5), 897–910. https://doi.org/10.1017/S0033291714001950

Guiraud, J. A., Tomalski, P., Kushnerenko, E., Ribeiro, H., Davies, K., Charman, T., Elsabbagh, M., & Johnson, M. H. (2012). Atypical audiovisual speech integration in infants at risk for autism. *PLoS ONE*, *7*(5), e36428. https://doi.org/10.1371/journal.pone.0036428

Haebig, E., Saffran, J. R., & Ellis, W. S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, *58*(11), 1251–1263. https://doi.org/10.1111/jcpp.12734

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3), e12381. https://doi.org/10.1111/desc.12381

Irwin, J. R., & Brancazio, L. (2014). Seeing to hear? Patterns of gaze to speaking faces in children with autism spectrum disorders. *Frontiers in Psychology*, *5*, 397. https://doi.org/10.3389/fpsyg.2014.00397

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011). Can children with autism spectrum disorders "hear" a speaking face? *Child Development*, *82*(5), 1397–1403. https://doi.org/10.1111/j.1467-8624.2011.01619.x

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the U.K. autism community. *Autism*, *20*(4), 442–462. https://doi.org/10.1177/1362361315588200

Kim, S. H., Paul, R., Tager-Flusberg, H., Lord, C., Volkmar, F. R., Paul, R., Rogers, S. J., & Pelphrey, K. A. (2014). Language and communication in autism. In F. R. Volkmar, S. J. Rogers, R. Paul, & K. A. Pelphrey (Eds.), *Handbook of autism and pervasive developmental disorders* (4th ed., pp. 230–262). Wiley.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to nonsocial contingencies rather than biological motion. *Nature*, *459*(7244), 257–261. https://doi.org/10.1038/nature07868

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, *7*(3), 361–381. https://doi.org/10.1016/S0163-6383(84)80050-8

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule–2nd ed. (ADOS-2)*. Western Psychological Corporation.

Magnée, M. J. C. M., De Gelder, B., Van Engeland, H., & Kemner, C. (2008). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *Journal of Child Psychology and Psychiatry*, *49*(9), 995–1000. https://doi.org/10.1111/j.1469-7610.2008.01902.x

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(7), 1349–1358. https://doi.org/10.1007/s10803-007-0521-y

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. https://doi.org/10.1007/s10803-005-0040-7

Munhall, K., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *15*(2), 133–137. https://doi.org/10.1111/j.0963-7214.2004.01502010.x

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. G. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, *7*, 1245. https://doi.org/10.3389/fpsyg.2016.01245

Pathak, M., Bennett, A., & Shui, A. M. (2019). Correlates of adaptive behavior profiles in a large cohort of children with autism: The autism speaks Autism Treatment Network registry data. *Autism*, *23*(1), 87–99. https://doi.org/10.1177/1362361317733113

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Righi, G., Tenenbaum, E. J., McCormick, C., Blossom, M., Amso, D., & Sheinkopf, S. J. (2018). Sensitivity to audio-visual synchrony and its relation to language abilities in children with and without ASD. *Autism Research*, *11*(4), 645–653. https://doi.org/10.1002/aur.1918

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, *18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter international performance scale*. Western Psychological Services.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised (ADI-R)*. Western Psychological Services.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, *48*(8), 813–821. https://doi.org/10.1111/j.1469-7610.2007.01766.x

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., Camarata, S., & Wallace, M. T. (2016). Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Research*, *9*(7), 720–738. https://doi.org/10.1002/aur.1566

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2014). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in Psychology*, *5*, 379. https://doi.org/10.3389/fpsyg.2014.00379

Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2018). The cascading influence of multisensory processing on speech perception in autism. *Autism*, *22*(5), 609–624. https://doi.org/10.1177/1362361317704413

Tager-Flusberg, H., Plesa Skwerer, D., Joseph, R. M., Brukilacchio, B., Decker, J., Eggleston, B., Meyer, S., & Yoder, A. (2017). Conducting research with minimally verbal participants with autism spectrum disorder. *Autism*, *21*(7), 852–861. https://doi.org/10.1177/1362361316654605

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

Thurm, A., Manwaring, S. S., Swineford, L., & Farmer, C. (2015). Longitudinal study of symptom severity and language in minimally verbal children with autism. *Journal of Child Psychology and Psychiatry*, *56*(1), 97–104. https://doi.org/10.1111/jcpp.12285

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports*. Advance online publication. https://doi.org/10.1038/srep21756

Vihman, M. (2014). *Phonological development: The first two years*. Wiley.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wodka, E. L., Mathy, P., & Kalb, L. (2013). Predictors of phrase and fluent speech in children with autism and severe language delay. *Pediatrics*, *131*(4), e1128–e1134. https://doi.org/10.1542/peds.2012-2221

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

Yoder, P., Watson, L. R., & Lambert, W. (2015). Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *45*(5), 1254–1270. https://doi.org/10.1007/s10803-014-2286-4

(*Appendices follow*)

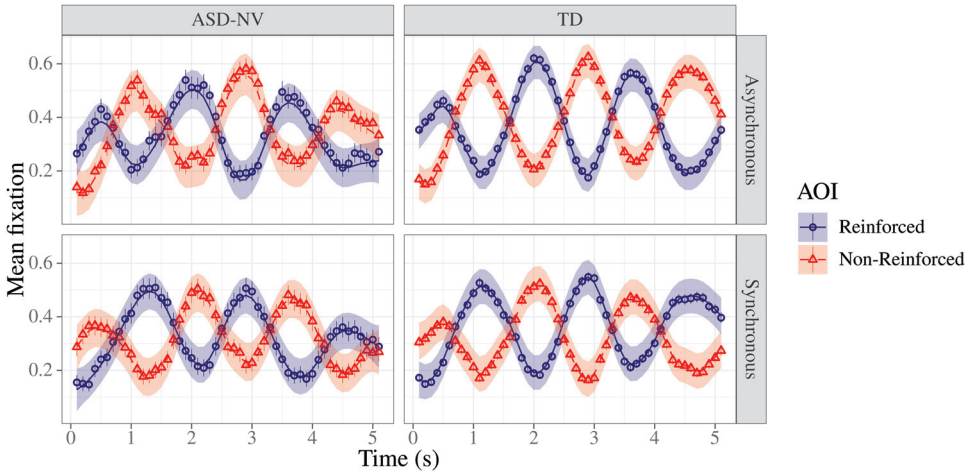
Appendix A

Excluding Verbal Autistic Children

The objective of our article is to gain better understanding of audio-visual integration in those autistic children who display little or no functional language. It is therefore important to determine the extent to which our results on audio-visual integration would be affected by the presence of some expressive language in a subset of our ASD group. For this reason, we subdivided all children in the ASD group in children with no expressive language at all (ASD-NV; speech condition: $n = 15$; nonsocial condition: $n = 21$) and those whose production included at least some words (ASD-V: speech condition: $n = 8$; nonsocial condition: $n = 5$). This clustering of the ASD group was grounded on the ADI-R criteria (item A30; Rutter et al., 2003) for defining nonverbal children, that is, the spontaneous and daily use of less than five functional words in the prior month. The appraisal of the child's functional speech were obtained through parents and educators reports, as well as by clinical observation during the testing sessions. We then implemented the same models as in the main text but excluding children in the ASD-V subgroup. Figures A1 and A3 display the mean values and the fitted curves of generalized additive models for the stimulus presentation phases of the two conditions; Figures A2 and A4 display the effects of categorical factors in the transition phases of the two conditions. As can be seen from these figures, the results reported in the main text remain essentially unaffected by the exclusion of verbal autistic children. The main difference is that once only nonverbal autistic children are kept in the analysis, the difference between the reinforced and the nonreinforced AOIs in the transition phase of the asynchronous version of the speech condition disappears (compare Figure 4 and Figure A2).

Figure A1

Speech Condition, Stimulus Presentation Phase

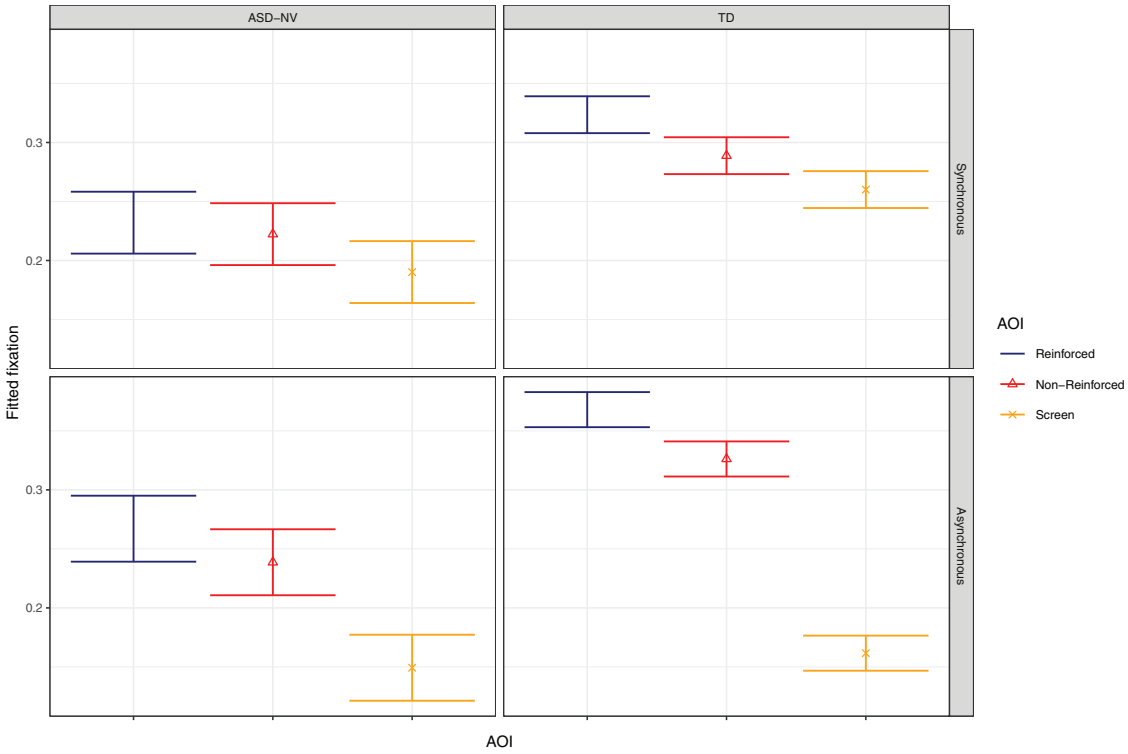


Note. Mean fixation values (per 100 ms bins) per AOI and fitted curves (random effects omitted); vertical bars represent standard errors of means and shadow ribbons represent 95% confidence intervals. ASD-NV = only nonverbal autistic children; TD = typically developing; AOI = area of interest. See the online article for the color version of this figure.

(*Appendices continue*)

Figure A2

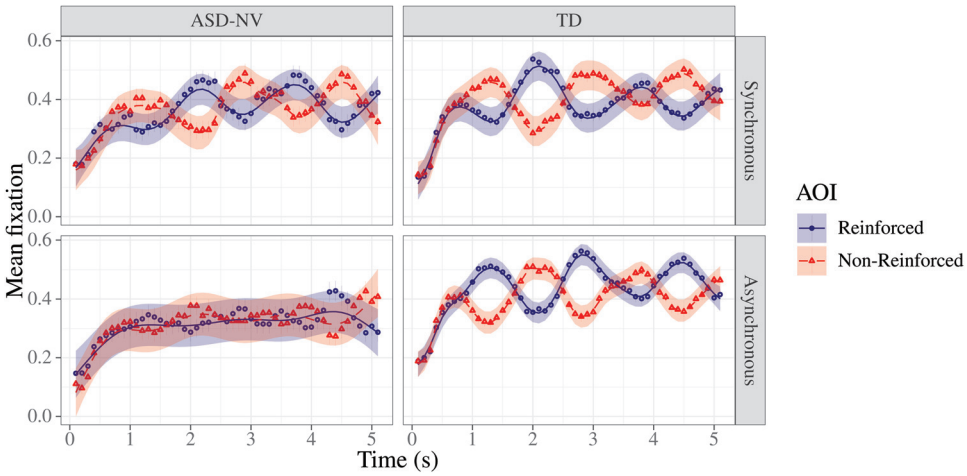
Speech Condition, Transition Phase; Excluding Verbal Autistic Children



Note. Effects of the categorical predictors (first 0.5 s); error bars represent 95% confidence intervals. ASD-NV = only nonverbal autistic children; TD = typically developing; AOI = area of interest. See the online article for the color version of this figure.

Figure A3

Nonsocial Condition, Stimulus Presentation Phase

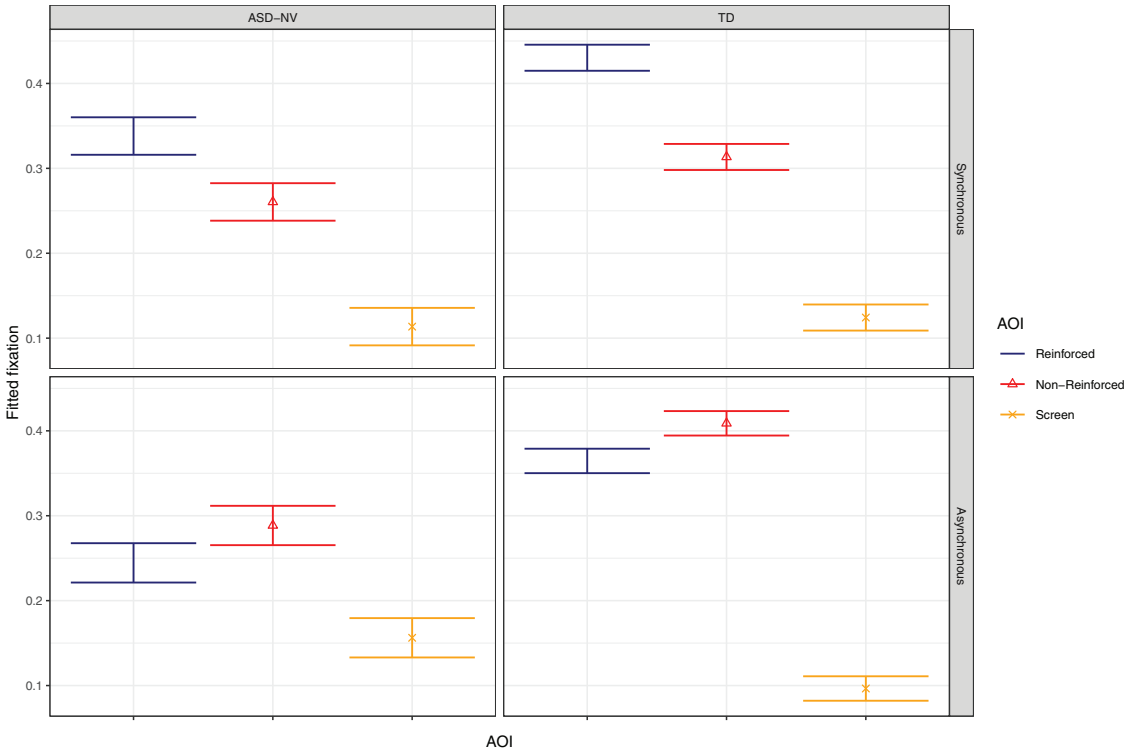


Note. Mean fixation values (per 100 ms bins) per AOI and fitted curves (random effects omitted); vertical bars represent standard errors of means and shadow ribbons represent 95% confidence intervals. ASD-NV = only nonverbal autistic children; TD = typically developing; AOI = area of interest. See the online article for the color version of this figure.

(*Appendices continue*)

Figure A4

Nonsocial Condition, Transition Phase; Excluding Verbal Autistic Children



Note. Effects of the categorical predictors (first 0.5 s); error bars represent 95% confidence intervals. ASD-NV = only nonverbal autistic children; TD = typically developing; AOI = area of interest. See the online article for the color version of this figure.

Appendix B

Supplementary Tables

Table B1

Speech Condition, Stimulus Presentation Phase. Summary of Generalized Additive Models; AOI Term is Contrast Coded

	ASD		TD	
Factor	Synch. version	Asynch. version	Synch. version	Asynch. version
	Parametric coefficient estimates (*SE*)			
Intercept	0.31*** (0.02)	0.35*** (0.02)	0.34*** (0.03)	0.38*** (0.03)
Nonreinforced AOI	−0.01*** ($0.36e^{-2}$)	0.03*** ($0.4e^{-2}$)	−0.03*** ($0.24e^{-2}$)	0.03*** ($0.23e^{-2}$)
	Smooth term estimated *df* (reference *df*)			
Time	16.02*** (19.70)	18.84*** (23.01)	21.81*** (26.34)	23.65*** (28.31)
Time: Nonreinforced AOI	19.41***	22.07*** (23.77)	25.24*** (30.04)	26.82*** (31.55)
Time by-participant	54.46*** (479)	43.93*** (439)	85.18*** (839)	90.37*** (919)
Time by-item	96.90*** (269)	87.11*** (269)	54.53*** (269)	174.19*** (269)

Note. ASD = autism spectrum disorder; TD = typically developing; AOI = area of interest.
*** $p < .001$.

(*Appendices continue*)

Table B2

Nonsocial Condition, Stimulus Presentation Phase. Summary of Generalized Additive Models; AOI Term is Contrast Coded

Factor	ASD		TD	
	Synch. version	Asynch. version	Synch. version	Asynch. version
	Parametric coefficient estimates *(SE)*			
Intercept	0.33*** (0.03)	0.35*** (0.02)	0.43*** (0.01)	0.38*** (0.02)
Nonreinforced AOI	$-0.18e^{-2}$ ($0.53e^{-2}$)	$-0.33e^{-2}$ ($0.54e^{-2}$)	-0.04*** ($0.35e^{-2}$)	0.02*** ($0.36e^{-2}$)
	Smooth term estimated *df* (reference *df)*			
Time	9.23*** (11.40)	11.40*** (14.09)	17.91*** (21.97)	18.11*** (22.17)
Time: Nonreinforced AOI	11.45*** (14.24)	15.14*** (18.74)	20.93*** (25.49)	19.57*** (23.93)
Time by-participant	59.49*** (519)	61.01*** (519)	79.40*** (959)	106.88*** (839)
Time by-item	57.84*** (269)	69.12*** (269)	30.77*** (269)	14** (269)

Note. ASD = autism spectrum disorder; TD = typically developing; AOI = area of interest.
** $p < .01$. *** $p < .001$.

Received November 20, 2018
Revision received December 16, 2020
Accepted December 28, 2020 ■