Infant Behavior and Development 78 (2025) 102026

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

# Sensitivity to temporal synchrony in audiovisual speech and language development in infants with an elevated likelihood of autism: A developmental review

Itziar Lozano [a,b,*,1], Ruth Campos [a,2], Mercedes Belinchón [a,3]

[a] Department of Basic Psychology, Faculty of Psychology, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, Iván Pavlov, 6, Madrid 28049, Spain

[b] Neurocognitive Development Lab, Institute of Psychology, Polish Academy of Sciences, Jaracza, 1, Warsaw 00-378, Poland

ARTICLE INFO

*Keywords:*
Infants with an elevated likelihood of autism
Temporal synchrony
Selective attention
Audio-visual speech
Language development

ABSTRACT

Detecting temporal synchrony in audiovisual speech in infancy is fundamental for socio-communicative development, especially for language acquisition. Autism is an early-onset and highly heritable neurodevelopmental condition often associated with language difficulties that usually extend to infants with an elevated likelihood of autism. Early susceptibilities in still unclear basic mechanisms may underlie these difficulties. Here, we discuss why sensitivity to temporal synchrony in audiovisual speech should be investigated in infants with an elevated likelihood of autism as a candidate mechanism underlying language difficulties. We then review direct and indirect eye-tracking evidence. Although scarce, some studies suggest that detection of temporal synchrony in audiovisual speech may be reduced in infant siblings (but evidence is mixed); however, this does not seem to account for language difficulties. Instead, a lack of relationship between selective attention to the articulating mouth and language development may be a plausible candidate mechanism. However, longitudinal studies tracking both sensitivity to temporal synchrony and selective attention to talking faces in the first year are needed for further clarification. Our discussion highlights gaps in the literature, future research directions and implications for domain-general approaches to the emergence of autism.

## 1. Introduction

### 1.1. *'What happens at the same time, goes together'*

Infants experience mostly multimodal events early in life, particularly in socio-communicative contexts (e.g., visuo-tactile stimulation). Audiovisual speaking faces are one of the most common multimodal events in adult-infant dyads, requiring infants to simultaneously hear speech sounds and see lip movements. Effective learning from this event requires the appropriate integration of audiovisual information (Wallace et al., 2019), a skill commonly referred to as *audiovisual processing*. As mature perceivers, adults are

* Correspondence to: Jaracza, 1, Warsaw 00-378, Poland
*E-mail addresses:* ilozano@psych.pan.pl (I. Lozano), ruth.campos@uam.es (R. Campos), mercedes.belinchon@gmail.com (M. Belinchón).
[1] ORCID: 0000–0003-4280–4747
[2] ORCID: 0000–0003-0209–6002
[3] ORCID: 0000–0001-5367–279X

https://doi.org/10.1016/j.infbeh.2024.102026

Available online 27 January 2025

capable of this perceptual endeavor, but how do infants face this challenge?

The Intersensory Redundancy Hypothesis (IRH, henceforth; Bahrick & Lickliter, 2002; 2014) has made three major theoretical contributions to this question. First, to overcome the challenge of audiovisual processing, infants rely on amodal properties of physical stimuli, such as rhythm or tempo, which are common to all sensory modalities. Among these, temporal synchrony detection would be the most important perceptual cue for organizing perceptual experience in early development, as it allows infants to bind auditory and visual stimuli into a single percept (Bahrick et al., 2004). Temporal synchrony is defined here as "changes in events that occur at the same moment in time" across different sensory modalities (Bahrick & Hollich, 2008, p. 164). Although this cue is permanently embedded in audiovisual information, infants do not show full sensitivity to it at birth (Lewkowicz, 1996; 2010, but see Lewkowicz et al., 2010); instead, it gradually improves as infants are exposed to multimodal experiences, especially in natural social interactions (Bahrick, 2010). Second, it hypothesizes how this ability develops. The detection of amodal properties would guide infants' attention, helping them to attend primarily to multimodal events (and, by extension, audiovisual events). This implies that the overlap of information in two sensory modalities is not problematic for infants, but helps, because redundancy makes information more salient. For example, 'faces become especially salient to infants because they are frequently encountered and typically the source of a great deal of intersensory redundancy' (Bahrick, 2010, p. 149). Importantly, 'redundancy' refers to the simultaneous availability and temporal synchronization of the same information across two or more sensory modalities (Bahrick & Lickliter, 2000), whereas 'saliency' refers to the 'attentional priority' that infants give to this information over other types of events (e.g., non-redundant as well as unimodal). Third, this approach highlights the key role of selective attention in the development of audiovisual processing, as the trajectory of changes in infants' selective attention to audiovisual redundancy is tightly coupled to changes in sensitivity to temporal synchrony and hence to changes in audiovisual processing (Lewkowicz, 2010).

Studies on the development of audiovisual speech processing in infancy have contributed to the above question in three ways. First, they have identified when sensitivity to temporal synchrony emerges in development. Despite mixed evidence (see the meta-analysis by Cox et al., 2022), the second half of the first year appears to be a crucial time-point (Hillairet de Boisferon et al., 2017; Pons & Lewkowicz, 2014-Experiment 1). Second, they found empirical support for the idea of the IRH that developmental changes in infants' selective attention affect their response to audiovisual inputs. In the context of audiovisual speech, selective attention refers to preferential attention to different internal facial features (i.e., the eyes and mouth). Infants' preference for the mouth (relative to the eyes) is not constant throughout the first year, but follows an inverted U-shaped pattern (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015): preference for the eyes at 4 months, a sharp increase in mouth-looking—the main source of audiovisual redundancy—at 8 months, and an equal preference for both areas at 12 months. Interestingly, the shift to the mouth coincides in timing with the onset of infants' detection of temporal asynchronies in audiovisual speech (i.e., 8 months; Pons & Lewkowicz, 2014). This suggests that the onset of temporal synchrony detection in talking faces may be closely intertwined with changes in infants' selective attention to the same event. Third, the type of infants' prior experience with audiovisual speech redundancy is likely to modulate their patterns of selective attention to talking faces (for a systematic review, see Bastianello et al., 2022).

For example, bilingual infants, who experience 'two types' of audiovisual redundancy, attend to the mouth for a longer period than monolinguals during development and maintain this preference at 12 months. In contrast, monolinguals show equivalent attention to the eyes and mouth, possibly suggesting that prolonged focus on the mouth may facilitate bilingual infants' acquisition of two languages (Pons et al., 2015). Relatedly, bimodal bilinguals (i.e., hearing infants with deaf mothers), who experience reduced audiovisual redundancy, do not shift attention to the mouth at 8 months (Mercure et al., 2019). These findings suggest that the type of early exposure to redundancy in audiovisual speech shapes infants' patterns of selective attention to this event, particularly during the first year.

### *1.2. Beyond audiovisual processing: The role of sensitivity to temporal synchrony in early development*

Sensitivity to temporal synchrony has been considered a domain-general ability (Lewkowicz, 2014; Pons et al., 2012). Although essential for audiovisual processing, its role in early development extends beyond the perceptual domain, supporting both infants' understanding of the physical world and the development of socio-communicative skills. This ability helps infants to cope with a changing environment by finding coherence, which facilitates multiple functions.

From 4 months, the detection of temporal synchrony supports object perception (Bahrick, 1983; Spelke, 1979). Infants link the visual and acoustic dimensions of an object by detecting the temporal synchrony between its sound and visual impact (Spelke, 1979). This ability also shapes socio-communicative skills during the first two years, both non-linguistic and linguistic. In the non-linguistic domain, the detection of temporal synchrony between faces and voices is necessary for infants' discrimination of affect as early as 4 months (Flom & Bahrick, 2007, Experiment 4). It also helps infants to engage in shared visual attention, an early form of joint attention (Flom et al., 2004), and to imitate facial expressions by coupling changes in lip movements with corresponding changes in spectro-temporal acoustic features (see Viswanathan et al., 2024 for a review). Linguistically, infants' sensitivity to temporal synchrony supports preverbal skills such as speech segmentation (Hollich et al., 2005), selective attention to lip movements and vocal sounds (Bahrick & Lickliter, 2002), learning word-object associations (Gogate et al., 2000), and enhancing word learning by directing infants' attention to synchronously named and gestured objects (de Villiers Rader & Zukow-Goldring, 2012).

Sensitivity to temporal synchrony contributes to early learning across domains, but is particularly important for the development of social skills. Given the demands of the social world, which is characterized by containing highly unpredictable, dynamic, and complex events related to people (e.g., emotions, gestures, biological motion, or audiovisual speech; Elsabbagh & Johnson, 2016), the detection of temporal synchrony seems particularly necessary for the efficient processing of social stimuli. It helps to create a more predictable environment and reduce perceptual uncertainty by identifying regularities. In contrast, detecting this cue may be less necessary for processing more regular events in non-social contexts, such as objects.

Audiovisual speech is arguably the most important social stimulus in infants' social lives (Soto-Faraco et al., 2012). It is unpredictable, and temporally complex, containing concurrent dynamic visual (i.e. facial articulatory movements) and auditory (i.e. vocalizations) information (Chandrasekaran et al., 2009). Given these characteristics, the detection of temporal synchrony helps infants to effectively process audiovisual speech during social interactions by making it more predictable, ultimately supporting later language development (Edgar et al., 2022).

## 2. When redundancy may not help: Difficulties in sensitivity to temporal synchrony in audiovisual speech in autism

Sensitivity to temporal synchrony and audiovisual processing may develop differently in several neurodevelopmental and clinical conditions, including Specific Language Impairment (Pons et al., 2013), prematurity (Imafuku et al., 2019), Down syndrome and Fragile X syndrome (D'Souza et al., 2016; Hill et al., 2012), and Autism Spectrum Condition (autism, hereafter) (Bebko et al., 2006; Righi et al., 2018).

Autism is a neurodevelopmental condition characterized by difficulties in social interaction and communication, and restricted and repetitive behaviors, interests, or activities (American Psychiatric Association, 2022). The latter dimension includes differences in sensory processing (e.g., expressed as hyper/hyposensitivity). Investigating potential difficulties in audiovisual integration and temporal asynchrony detection in autism rather than other neurodevelopmental conditions has been of interest to test whether potential early difficulties in these two perceptual abilities may underlie the core features of social functioning that define autism (Tager-Flusberg, 2010). As these mechanisms are key to the development of typical socio-communicative skills, investigating them in autism helps to examine potentially altered relationships between the development of difficulties in early low-level perceptual skills and later higher-order socio-communicative skills. These relations are of interest for both non-linguistic and linguistic social skills, given that difficulties in language development are common from early in life in most autistic individuals (Eigsti et al., 2011; Miniscalco et al., 2012; Tager-Flusberg et al., 2005), although they are highly heterogeneous (Schaeffer et al., 2023) and are not a core feature of the condition.

### 2.1. *Infants with an elevated likelihood of autism: Investigating potential difficulties as they emerge*

Autism is highly heritable (Bai et al., 2019). Infant siblings of children with an autism diagnosis have an increased likelihood of developing the condition themselves compared to the neurotypical population. Approximately 19 % of infants with an elevated likelihood of autism (henceforth, EL-infants) meet the criteria for the condition at the age of diagnosis (Ozonoff et al., 2011), whereas the prevalence in the general population is ~1 % (Zeidan et al., 2022). More importantly, approximately 11 % of EL-infants without autism show mild to moderate developmental delays and adaptive functioning difficulties at 3 years of age (Charman et al., 2017), sometimes observed even as early as their first birthday. Over the past two decades, prospective longitudinal studies of infant siblings have reported that EL-infants without autism show developmental difficulties in several domains, including cognitive, motor, language, and social (Ozonoff et al., 2014), compared to infants with typical autism likelihood (i.e., infants whose older sibling does not have a diagnosis of autism; hereafter TL-infants). In addition, a subgroup of these infants show subclinical autistic traits at 3 years of age (Messinger et al., 2013).

Other prospective longitudinal studies have examined whether EL-infants differ from TL-infants at the group level in several foundational developmental skills. To do this, these studies have followed infant siblings from early in life across various domains, comparing groups based on autism likelihood rather than diagnostic outcomes. Overall, the findings suggest that, at the group level, EL-infants differ from TL-infants on several behavioral, cognitive, and neurobiological characteristics observed in autistic individuals, especially in the socio-communicative domain, but not exclusively. For example, EL-infants differ from TL-infants in attentional disengagement (Elsabbagh et al., 2009a), visual orientation (Elsabbagh et al., 2013b), direct gaze processing (Elsabbagh et al., 2009b), face processing (Key & Stone, 2012), speech processing (Guiraud et al., 2012), and postural control (Bhat et al., 2011). Importantly, despite heterogeneity, most group differences emerge around the second half of the first year (see Elsabbagh & Johnson, 2010 and Tager-Flusberg, 2010 for reviews). These findings suggest that potential group-level difficulties in EL-infants may extend beyond abilities related to the core autistic traits, also including domain-general difficulties in perceptual, attentional, and motor abilities that are mechanisms underlying higher-level difficulties before they potentially emerge. Therefore, even if they do not develop the condition themselves, EL-infants as a group may share milder cognitive difficulties and thus develop differently from TL-infants (Yirmiya et al., 2006).

### 2.2. *Two candidate mechanisms of language difficulties in infant siblings?*

Socio-communicative abilities encompass a wide range of verbal and non-verbal skills necessary for reciprocal social interactions, including language. The role of language in autism is core, as early language skills are one of the best predictors of later adaptative functioning of autistic individuals (Szatmari, et al., 2003), and language delay is one of the first parental concerns in autistic toddlers (Herlihy et al., 2015). Although not all autistic individuals show language difficulties (Tager-Flusberg et al., 2005) and there is high variability, they are often observed across the lifespan (Eigsti et al., 2011). For example, autistic toddlers in fact show reduced speech vocalizations between 18 and 24 months (Plumb & Wetherby, 2013) and, later in childhood, different profiles of receptive versus expressive language skills compared to their neurotypical peers (Miniscalco et al., 2012). Language difficulties also persist into adulthood (Howlin, 2003).

Given the high heritability of autism, the language difficulties associated with the autism phenotype often extend to first-degree relatives. For example, parents of autistic children exhibit pragmatic language difficulties (Whitehouse et al., 2007). EL-infants also show difficulties in language development at the group level (Belteki et al., 2022; Hudry et al., 2014), such as lower consonant production, fewer speech-like vocalizations (Paul et al., 2011), later onset of babbling, and delayed language production and comprehension in toddlerhood (Iverson & Wozniak, 2007).

In line with other authors in the field (Bahrick, 2010; Bahrick & Todd, 2012), we argue that the language difficulties observed in autistic individuals across the lifespan and in their relatives—including infant siblings—may be expressed earlier as difficulties in the perceptual *mechanisms* that contribute to language acquisition as precursors, with two main candidates being potentially affected.

#### 2.2.1. Candidate 1: Sensitivity to temporal synchrony in audiovisual speech

As argued above, sensitivity to temporal synchrony in neurotypical infants is crucial for early language acquisition (Cox et al., 2022). Given its role in audiovisual integration, especially in *audiovisual speech*, its detection is likely to be an important mechanism underlying the early typical development of preverbal and verbal skills (e.g., vocal imitation, babbling, speech segmentation, and word learning; e.g., Hollich et al., 2005; Gogate et al., 2000). In support of this functional link, in the neurotypical population, temporal asynchrony detection in *audiovisual speech* is positively associated with receptive and expressive language in childhood (Righi et al., 2018). Given the well-documented difficulties in language development at the group level in EL-infants (Belteki et al., 2022; Hudry et al., 2014; Paul et al., 2011), difficulties in temporal synchrony detection in *audiovisual speech* may be present in this group from the earliest developmental time-points, before potential language difficulties emerge.

#### 2.2.2. Candidate 2: Selective attention in audiovisual speech

Potential difficulties may also be present in EL-infants in selective attention in *audiovisual speech,* another crucial low-level mechanism for language acquisition (Bahrick, 2010; Bahrick & Todd, 2012). Selective attention involves focusing on certain elements or events while ignoring others (Bahrick & Lickliter, 2014). In the context of *audiovisual speech* processing, it refers to the preferential attention to different facial features (i.e., the eyes and mouth). As selective attention provides the basis for what is perceived, learned, and remembered in typical development, it also represents an essential ability to be investigated in autism and, more broadly, in EL-infants (Bahrick, 2010; Bahrick & Todd, 2012). As Walker-Andrews (1994) quoted from J.J. Gibson's words (Gibson, 1966), 'learning of new meaning is an education of attention rather than an accrual of associations'.Although temporal synchrony detection is considered to be a more fundamental mechanism than selective attention to talking faces for the development of language and other sociocommunicative skills, both mechanisms would develop in a closely intertwined manner from the first months of life (Bahrick, 2010). Therefore, although they are separate candidate mechanisms for potential language difficulties, we believe that they should be investigated together.

We argue that investigating selective attention in the early development of EL-infants, particularly in *audiovisual speech*, is relevant for at least two reasons.

The first reason is that selective attention modulates infants' perceptual experience with audiovisual redundancy and, thus, temporal synchrony detection. As we pointed out in Section 1, neurotypical infants change their selective attention to talking faces during the first year, which affects their access to audiovisual redundancy (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). Infants do not prioritize mouth-looking equally across development, suggesting that audiovisual redundancy may not always be equally helpful to them. Furthermore, mouth-looking in early infancy may not only support the detection of temporal synchrony in *audiovisual speech*, but may itself be an important mechanism for language development (see Edgar et al., 2023 for longitudinal evidence). For example, increased mouth-looking during the second half of the first year concurrently and prospectively predicts receptive and expressive language (Kushnerenko et al., 2013; Lozano et al., 2022; Tsang et al., 2018; Young et al., 2009). Given the group-level language difficulties observed in EL-infants (Belteki et al., 2022; Hudry et al., 2014), we propose that it is of interest to explore whether they differ from TL-infants in their changes in selective attention to *audiovisual speech* over the first year of life. If so, a different developmental trajectory in this ability may reduce their attentional access to relevant visual cues in *audiovisual speech* on which typical language acquisition relies. Similarly, difficulties in efficiently attending to a speaking face may also reduce EL-infants' opportunities to learn to detect temporal synchrony, which involves an experience-dependent process of perceptual learning (Gibson & Pick, 2000). Ultimately, this may reduce the fidelity with which they process linguistic experiences.

The second reason is that we believe that selective attention in *audiovisual speech* processing is a top-down mechanism. This means that infants' attention to specific features of a speaking face is driven by their prior experience with that event. As noted in Section 1, research with neurotypical bilingual and bimodal bilingual infants suggests that experience with *audiovisual speech* modulates how salient redundancy is to infants (Pons et al., 2015; Mercure et al., 2019). EL-infants are more likely than TL-infants to have different experiences that are constrained by both neurocognitive differences and their interactions with their environment (Campos, 2018; Johnson et al., 2015). Therefore, investigating whether EL-infants show different changes in selective attention to *audiovisual speaking faces* compared to their neurotypical peers may provide insights into their ability to learn from social contexts. Learning from social events is crucial for the development of socio-communicative skills, including language acquisition, which often develops differently in EL-infants.

## 3. Aim of this review

This review investigates whether sensitivity to temporal synchrony in *audiovisual speech* plays a role in accounting for potential difficulties in language development in infants with an elevated likelihood of autism. As argued aboved, we focus on audio-visual speech, among all the possible sensory modalities and events, because of its crucial role in the typical acquisition of language and other socio-communicative skills, and the core social features and related language difficulties that define autism. To address this question, we review and critically discuss findings from prospective longitudinal eye-tracking studies in this population. Eye-tracking is an advanced method that allows the characterization of fine-grained attentional patterns to audiovisual speech in young developmental populations. In addition, it has the potential to identify endophenotypes in infant siblings at the behavioral level, linking their familial autism likelihood to later developmental outcomes (Nayar et al., 2022), such as the two candidate mechanisms we propose in this review.

We begin by reviewing *direct evidence* on sensitivity to temporal synchrony in audiovisual speech in infant siblings. Next, we review *indirect evidence* from studies of audiovisual speech integration and visual attention to talking faces during the first year of life in this population. Following each set of studies, we briefly discuss gaps in the literature that could be addressed by future research. Finally, we provide a general discussion of these findings, highlighting future challenges and theoretical implications for domain-general approaches to the emergence of autism, in particular the *Intersensory Impairment Hypothesis* (Bahrick & Todd, 2012) and Neuroconstructivism (Karmiloff-Smith, 1998). We conclude by proposing directions for future research.

Although this is a narrative review, we briefly report our search strategy for transparency (for a comprehensive systematic review, see Mastergeorge et al., 2021). Eye-tracking studies were identified through searches in PubMed, Web of Science, and Google Scholar between May and June 2024. A detailed summary of the keywords and Boolean operators we used to search for both direct and indirect evidence in each database is provided in the Supplementary Materials (Table S1).

## 4. Direct and indirect evidence on sensitivity to temporal synchrony in audiovisual speech in infant siblings

### 4.1. Direct evidence

To our knowledge, only one study has *directly tested* whether EL-infants show difficulties in their sensitivity to temporal synchrony in the context of *audiovisual speech* and, if so, whether this predicts later language difficulties (Suri et al., 2023). Using the habituation-test paradigm, the authors assessed EL (n = 35) and TL (n = 53) infants aged 4–24 months in their sensitivity to audiovisual temporal asynchrony for social (speaking face uttering syllables) and non-social (bouncing ball) events, and their relations to language productive outcomes in toddlerhood (at 17–30 months). EL-infants were less sensitive to detecting audiovisual asynchronies than their neurotypical peers. Crucially, compared to TL-infants, EL-infants showed a significantly larger temporal binding window (i.e., the range of asynchronies under which an auditory and a visual event are likely to be bound into a unified percept was wider than in TL peers, detecting mean temporal asynchronies of 671 ms vs. 575 ms, respectively) only in the *audiovisual speaking* social event, but the groups did not differ in the non-social event. Moreover, in TL-infants, sensitivity to temporal asynchronies in either event did not predict vocabulary production in toddlerhood. Conversely, in EL-infants, a wider temporal binding window only in the audiovisual speech event predicted a larger productive vocabulary.

These results suggest that sensitivity to temporal synchrony appears to be reduced in EL-infants specifically in the *audiovisual speech* domain, reflecting potential difficulties at the group level. However, they do not support the idea that reduced detection of temporal asynchrony underlies the development of language difficulties in infants' siblings. Suri et al. interpreted their findings as consistent with previous studies in autistic children that have shown greater difficulties in detecting temporal asynchronies in *audiovisual speech* than in non-social events (Bebko et al., 2006; Irwin et al., 2011; Mongillo et al., 2008; Stevenson et al., 2014); but as inconsistent with the idea that a smaller temporal binding window reflects greater sensitivity to audiovisual asynchrony and better integration of auditory and visual speech cues (Zerr et al., 2019). This was so because TL infants—who exhibited a smaller temporal binding window than EL infants—showed no relation between sensitivity to temporal synchrony and later vocabulary in either event. The authors also speculated that the unexpected association between a larger temporal binding window and language outcomes may suggest that EL-infants may use different strategies to acquire language.

While seminal, this study has several limitations, most of which are acknowledged by the authors.

First, the lack of eye-tracking technology prevents the authors from ruling out the possibility that EL-infants selectively attended to the speaking face, particularly to the mouth, differently from neurotypical infants (see Lozano et al., 2024). As Todd and Bahrick (2022) observed in autistic children (2–5 years old), it may be that synchrony detection in *audiovisual speaking faces* in EL-infants is comparable to that of neurotypical peers and is beneficial for language acquisition, but there is a lack of relation between this perceptual ability and selective attention to the mouth of talking faces. In contrast, in neurotypical infants, both skills would be intertwined to benefit language outcomes. Alternatively, EL-infants may compensate for reduced temporal asynchrony sensitivity in *audiovisual speech* by paying more attention to the mouth than neurotypical infants. Future studies should use eye-tracking to test these two possibilities by simultaneously measuring sensitivity to temporal synchrony and selective attention to the eyes and mouth within the same EL-infants, ideally using separate tasks.

Second, the authors did not compare EL-infants with and without autism with TL-infants, which is necessary to clarify whether the difficulties they found in sensitivity to temporal synchrony in *audiovisual speech* were driven by a subset of EL-infants with a later diagnosis of autism, or are present in EL-infants as a whole. Therefore, it remains unknown whether difficulties in this mechanism may reflect an early subclinical trait linked to a heightened familial likelihood of autism (potentially part of the infant's broader autism phenotype; Yirmiya & Ozonoff, 2007) or an early autism marker. Thus, future infant sibling studies should assess autism outcomes and make group comparisons based on both autism likelihood and diagnostic outcomes. It is also necessary to test within-group associations (in EL-infants as a group, in EL-infants with and without autism separately, relative to TL-infants) between temporal synchrony detection and later language outcomes and the severity of autistic traits. This would help to clarify whether difficulties in this perceptual ability are a defining feature of autism or rather a risk factor for the language difficulties associated with the condition.

Third, Suri and collaborators tested EL-infants across a wide age range (4–24 months) and assessed language outcomes at different ages for each infant (17–30 months), which may have hindered robust identification of the onset of difficulties in sensitivity to temporal synchrony. This heterogeneity may also have obscured potential differences in the associations between this mechanism and language development in this population, as studies with neurotypical infants suggest that the time when it may be more important for later language support is the second half of the first year (Pons & Lewkowicz, 2014). Future studies with infant siblings should test this perceptual ability at more specific and functionally relevant time-points in the first year—6–12 months—and assess concurrent and prospective language outcomes at the same age in all infants.

Fourth, the stimuli used—syllables—were not as close to infants' natural social experiences as audiovisual *fluent speech.* Different associations between sensitivity to temporal synchrony and later difficulties in language outcomes in EL-infants may only be observable in more complex naturalistic settings closer to EL-infants' linguistic experiences, as this may increase the task's demands. Future research with EL-infants may benefit from the use of *audiovisual fluent speech,* ideally spoken by real people during dyadic caregiver-infant interaction in a 'live' eye-tracking setting (Birulés et al., 2023), to take a more naturalistic approach. Finally, the inclusion of preterm and low-birth-weight infants in the EL-infant group may have introduced additional heterogeneity, obscuring potentially affected associations between reduced temporal synchrony detection and language outcomes.

More recently, in a preliminary longitudinal study, Lozano et al. (2024) investigated whether EL-infants (n = 14) at 4, 8, and 12 months showed difficulties in temporal synchrony detection and selective attention to *audiovisual fluent speech* compared to neurotypical peers (n = 15). No evidence of group differences in detecting temporal synchrony was found, with both groups succeeding at 12 months. However, the authors found preliminary evidence of reduced mouth-looking in EL-infants compared to TL-infants only at 12 months. Nonetheless, due to the small sample size, the authors did not make group comparisons by autism outcome or analyze the associations between these two mechanisms and language development and autism diagnostic outcomes, so no conclusions could be drawn about these functional relations and their specificity to infant siblings diagnosed with autism or at the group level.

### 4.2. *Indirect evidence*

Beyond the studies by Suri et al. (2023) and Lozano et al. (2024), to our knowledge, there is no other direct published evidence. However, *indirect evidence* may also contribute to the aim of this review. In this section, we first review literature evaluating potential difficulties in EL-infants in audiovisual integration of speech (Bahrick & Todd, 2012), a perceptual ability to which sensitivity to temporal synchrony directly contributes. We then present indirect evidence from eye-tracking studies testing EL-infants with audiovisual stimuli of faces and speech. Although not designed to examine temporal synchrony detection, these studies are collectively informative of whether EL-infants may show visual attentional difficulties when speech and faces are presented redundantly. We conclude this subsection by presenting the few studies that have investigated the relationship between mouth looking in infancy and language outcomes in the second year of life, and by discussing the overall findings of this set of indirect studies.

#### 4.2.1. *Audiovisual speech integration studies*

At least one study suggests that infant siblings may have a reduced ability to integrate audiovisual information in the context of speech. Audiovisual speech integration has classically been measured through the McGurk effect (McGurk & Macdonald, 1976), a perceptual phenomenon in which visual speech information (mouth articulatory movements) influences the perception of speech sounds. In an eye-tracking study, Guiraud et al. (2012) used the McGurk paradigm to test this possibility in 9-month-old EL-infants (n = 26; n = 17 neurotypical). In a preferential-looking paradigm, infants were shown with a congruent pair (auditory /ga/ paired with visual /ga/, or auditory /ba/ paired with visual /ba/) and an incongruent pair. The incongruent pair consisted of either an auditory /ba/ syllable paired with a video of a face articulating /ga/ (commonly perceived as a 'fusible' percept /da/, the McGurk effect), or the reverse pairing (a face articulating /ba/ paired with an auditory /ga/), which is commonly perceived as /bga/ (a mismatch, a non-fusible effect). In contrast to their TL peers, EL-infants did not perceive the McGurk illusion, which is usually considered an indicator of successful audiovisual integration. As this is also seen in autistic children (Taylor et al., 2010), the authors interpreted this result as suggesting that infant siblings, as a group, may share difficulties in audiovisual speech integration.

This finding has been further supported at the neurophysiological level. An ERP (event-related potential) study (Riva et al., 2022) documented a reduced audiovisual speech integration in EL-infants at 12 months. In a McGurk paradigm, TL-infants (n = 19) showed larger mean amplitude responses over the left temporal area to the audiovisual mismatch condition than to the audiovisual matching condition, whereas EL-infants (n = 21) showed equivalent ERP responses between conditions. This suggests that EL-infants, as a group, do not perceive incongruencies between auditory and visual *speech* stimuli in a comparable manner to their neurotypical peers, supporting the prediction of difficulties in audiovisual speech integration in this population. As there were no comparisons based on autism diagnostic outcomes, it remains possible that the subgroup of EL-infants with later autism drove the group effects. Furthermore, associations with language outcomes were not examined as this was outside the scope of the study, so it is inconclusive whether the audiovisual integration difficulties observed by Riva et al. cascade into later language difficulties in EL-infants. Interestingly, however, the authors did find an association between reduced audiovisual speech integration in talking faces and greater hypo-responsiveness to sensory input. As suggested by Riva and collaborators, this relation should be further investigated to test the potential cascading effects on social communication. Infants who are unresponsive to relevant sensory events in their environment may lose opportunities for social interaction, which is crucial for socio-communicative development in general and, we would add, for language acquisition in particular.

Taken together, this indirect evidence, when interpreted in the context of the language difficulties found in EL-infants as a group

(Belteki et al., 2022; Hudry et al., 2014) and the well-established functional link between audiovisual speech processing and language outcomes in neurotypical infants (Edgar et al., 2022; Kushnerenko et al., 2013), only partially supports the idea that difficulties in audiovisual processing may precede difficulties in language development. According to the findings of Guiraud et al. (2012) and Riva et al. (2022), difficulties in audiovisual speech processing seem to emerge at the group level in EL-infants towards the end of the first year (at 9 and 12 months). However, these studies did not test associations between this difficulty and potential future language outcomes or the severity of autistic traits, so the potential cascading effects remain unclear. Furthermore, the group comparisons were again based on autism likelihood, not diagnostic status, so it remains unclear whether the audiovisual speech processing difficulties observed in the whole group were driven by a subset of infants' siblings with autism. Theoretically, the primary difficulty that precedes and cascades into difficulties in audiovisual speech integration would be in temporal synchrony detection (Bahrick & Todd, 2012). Therefore, future behavioral and neural studies of audiovisual speech integration in infant siblings should measure—ideally within the same participants and using the same event—their performance in detecting temporal synchrony, and test within-group associations between these two perceptual abilities and their predictive role in the development of later language outcomes and autistic traits.

#### *4.2.2. Studies exploring visual attention to audiovisual speech*

A systematic search of the eye-tracking research conducted with EL-infants over the last decade reveals that several studies have investigated social competence (or 'social attention'; see Falck-Ytter et al., 2022, for a review of the complexity of this umbrella term) using tasks containing audiovisual speech stimuli. Despite the heterogeneity of these studies (e.g., in terms of stimulus content, age of testing, or comparison groups—see Table S2 for methodological details), they share some characteristics: (1) they generally aim to test which gaze patterns EL-infants perform when exploring social information (i.e., events related to or involving interaction with people, although not all stimuli used fit this definition), (2) at early developmental ages (6–24 months), (3) they compare EL-infants with at least one group of TL-infants, sometimes further classifying and comparing EL-infants based on diagnostic outcomes (usually, EL-autism, EL-atypical, EL-typical), (4) by showing infants audiovisual stimuli (often social, mainly female speaking faces), (5) that vary in complexity (from isolated faces to social scenes).

Complexity here usually refers to the manipulation of several dimensions of the stimuli, such as dynamic/static, or level of predictability/variability (i.e., in the sense of rapidly changing over time; e.g., Blasi et al., 2013). It is generally assumed that the more dynamic and less predictable a social stimulus is, the more complex it is. Finally, 6) they usually measure EL-infants' visual attention patterns when exploring audiovisual information, mainly their preference for facial features (i.e., the eyes and mouth). Thus, although indirect, this research may provide meaningful evidence on how EL-infants' attentional performance functions when dealing with two sensory modalities (auditory and visual). Notably, we excluded studies that used social audiovisual stimuli without realistic people (e.g., biological motion; Falck-Ytter et al., 2018).

*4.2.2.1. Evidence for group differences.* In a prospective study, Shic et al. (2014) investigated whether preference for social information—a bias that neurotypical infants show from birth—works differently in EL-infants with different diagnostic outcomes (EL-with autism, EL-atypical, or EL-typical) at 6 months. The task included social stimuli of increasing complexity: a still picture, a silent video with a moving smiling face, and an audiovisual clip with speech. Notably, only in the audiovisual condition (dynamic speech), did the subgroup of EL-infants later diagnosed with autism (but not the EL-infants as a whole) show significantly lower fixation times on the inner facial features and higher fixation times on the outer ones. Because auditory and visual information were only simultaneously available in this condition (and, thus, had to be integrated by the infants to succeed in the task), the authors concluded that "*the presence of speech might uniquely disturb the attention of infants who later develop autism at a critical developmental point when other infants are acquiring language and learning about the social world*" (Shic et al., 2014, p. 231).

This study does not allow us to examine whether EL-infants with autism have difficulties in audiovisual processing. This would have required the inclusion of a unimodal auditory speech condition. However, finding that the addition of speech to the task modulated infants' face scanning only in this subgroup suggests that the increased complexity of social stimuli (compared to a still face picture and a silent clip of a smiling face) differentially affects the patterns of selective attention to faces in EL-infants later diagnosed with autism. Consistently, other authors have also interpreted the results of this study as an indication that audiovisual speech introduces additional task demands, as it requires high temporal resolution integration (e.g., Elsabbagh & Johnson, 2016). This interpretation is consistent with the idea that EL-infants with autism may *adapt* their behavior to their processing abilities by withdrawing their attention from complex and unpredictable events.

Consistent with this, a seminal prospective longitudinal study conducted at earlier ages (2–24 months) found group differences in face scanning patterns (Jones & Klin, 2013). Infants watched an audiovisual clip of a female actress emulating a natural interaction, and their diagnostic outcomes were measured at 36 months. EL-infants who were later diagnosed with autism showed a mean decline in looking at the eyes from 2 to 6 months, unlike TL-infants. In contrast, there were no group differences in mouth-looking. Thus, compared to their neurotypical peers, EL-infants with autism showed a different pattern of visual fixation to the eyes of audiovisual speaking faces, following a different developmental trajectory already in the first half of the first year of life. This different attentional pattern was interpreted by the authors as being specific to EL-infants with autism outcomes, rather than to the group as a whole, suggesting that it may be a potential early marker of autism. However, as the group comparisons were based only on autism diagnostic outcomes, we argue that it cannot be ruled out that early difficulties in eyes-looking may extend to the whole group of EL-infants. Of note, only male participants were included in the analyses, which may limit the generalizability of these findings to female EL and neurotypical infants, as sex differences in attention to static and dynamic faces and preferences for the eyes and mouth have been documented in both populations (Chawarska et al., 2016; Kleberg et al., 2019; Lozano et al., 2022).

*4.2.2.2. No evidence of group differences.* However, other studies have found null results for group differences. In another prospective study, Chawarska et al. (2013) also assessed the visual preference of infants' siblings in a naturalistic social scene at 6 months of age. They used a task in which one condition consisted of a clip simulating a dyad with the participant, including a woman performing different types of activities while talking to the infant about scene elements. Consistent with the findings of Shic et al. (2014), infants who were later diagnosed with autism attended less to the social scene, as they spent a significantly lower proportion of fixation time looking at the actress, particularly her face. However, there were no group differences in the distribution of attention to the mouth or eyes. All infants, regardless of diagnostic outcome, spent more time looking at the eyes and mouth in the dyadic condition (and another 'shared attention' condition) than in the conditions in which the actress was not trying to engage with the infant but performing activities.

Elsabbagh et al. (2014) also reported no differences between 7- and 14-month-old EL-infants with autism and TL-infants in their visual scanning strategies or in the distribution of attention between the eyes and mouth. This suggests that the scanning of complex social scenes does not differ significantly between infants who go on to develop autism and those who do not. This was also replicated when the groups were compared on the basis of autism likelihood status. Interestingly, the task involved simple or complex dynamic scenes with multiple communicative signals (i.e., clips of a face moving its mouth, eyes, or hands as simpler conditions, and a clip of a *peek-a-boo* scene as the most complex condition) that did not include speech. Considering these results together with those of Shic et al. (2014), the presence or absence of speech in the task may explain the mixed evidence in the face scanning patterns of infants' siblings. While the use of audiovisual speech elicited differences in visual attention to internal facial features between EL-infants with autism and TL-infants (Shic et al., 2014), no group differences were found in its absence (Elsabbagh et al., 2014).

*4.2.2.3. Relations between mouth-looking and language outcomes.* Finally, only two published studies have measured mouth-looking during audiovisual speech events in infant siblings during infancy and related their performance to language outcomes in the second year of life.

In a prospective longitudinal study, Chawarska et al. (2022) compared EL-infants without autism with neurotypical infants at 12 months in their preferences for the face and mouth of a dynamic talking face uttering audiovisual speech. This was interspersed with non-speech episodes in which the actress was making a sandwich or looking at toys. Language outcomes at 18 months were also assessed using a developmental scale (Mullen Scales of Early Learning; MSEL, Mullen, 1995) and autism diagnostic outcomes and trait severity using a standardized scale (Autism Diagnostic Observation Schedule 2–Toddler Module; ADOS 2; Lord et al., 2012) at 18, 24, and 36 months. Importantly, EL-infants diagnosed with autism were excluded from the analyses. The authors found that EL-infants without autism did not differ from TL-infants in attention to the face or preference to the mouth. However, only TL-infants showed positive associations between attention to the face and mouth-looking at 12 months, and language outcomes at 18 months. This relation was absent in EL-infants without autism, suggesting that the functional link between increased mouth-looking and better language outcomes may potentially be different at the group level in infant siblings. This supports the idea that difficulties in selective attention to audiovisual speech may underlie difficulties in language development in emerging autism (Bahrick, 2010; Bahrick & Todd, 2012).

However, the lack of group comparisons based on autism diagnosis rather than autism likelihood prevents concluding whether this difficulty is also present in infants later diagnosed with autism. This is plausible, as Habayeb et al. (2021) found in a cross-sectional study that looking at the mouth was positively associated with expressive language in 10–25-month-old neurotypical toddlers and those with an autism diagnosis who had acquired their first words, compared with those who had not. In addition, autistic toddlers showed less mouth-looking than neurotypical peers. These findings suggest that, as Chawarska et al. (2022) point out, selective attention to audiovisual speech may be a potential endophenotype candidate mechanism that 'may lay between the distal genetic risk factors for autism and proximal overt expressions of these factors in behavior (i.e., language development difficulties)' (p. 1473).

Nevertheless, in contrast to Chawarska et al. (2022), Santapuram et al. (2022) found in a cross-sectional study no differences between EL-infants and TL-infants (6–18 months) in selective attention to the eyes and the mouth, and no direct associations between mouth-looking and concurrent expressive language regardless of likelihood status. Instead, they found that increased mouth-looking was indirectly related to better expressive language via increased joint engagement with the caregiver and prelinguistic vocal complexity in both groups. Unlike Chawarska et al.'s, these findings do not support the view that difficulties in selective attention to audiovisual speech may underlie difficulties in language development in infant siblings (Bahrick, 2010; Bahrick & Todd, 2012). Of note, the wide range of ages tested may have masked potential difficulties in selective attention to the mouth in audiovisual speech in EL-infants at the end of the first year, and their cascading effects on other preverbal communicative skills that may underlie potential language difficulties in autism.

*4.2.2.4. Differences in the candidates or the functional links to language?.* Albeit indirectly, the reviewed studies indicate that, at the group level, EL-infants show similar visual fixation patterns to neurotypical infants when scanning social information from both visual and auditory modalities (i.e., audiovisual speaking faces). This challenges the hypothesis that infant siblings, as a whole, show difficulties in processing audiovisual redundant speech events, as none of the eight indirect studies reviewed (summarized in Table S2) found group differences between EL and TL-infants in selective attention to audiovisual stimuli. However, in one of the only two studies that assessed language outcomes (Chawarska et al., 2022), the functional association between mouth-looking and expressive and receptive language was absent, in contrast to the positive associations found in neurotypical infants. This suggests that while the attentional ability itself may not be different in infant siblings, the functional link may be different compared to their neurotypical peers (but see Santapuram et al., 2022 for null results). Thus, selective attention to audiovisual speech may account for the language difficulties associated with an increased likelihood of autism.

Although not convergent, the reviewed studies also support the possibility that EL-infants with a diagnosis of autism do not scan relevant facial features similarly to neurotypical infants when *audiovisual speech* is involved, as two of the eight indirect studies reviewed (Shic et al., 2014; Jones & Klin, 2013) found difficulties in selective attention to audiovisual speech from 2 to 6 months specifically in infants with a later diagnosis of autism. In line with this, Habayeb et al. (2021) reported that toddlers with a diagnosis of autism showed reduced looking at the mouth compared to their neurotypical peers.

Although we cautiously view this research as indirect, given that infants' selective attention to talking faces appears to be closely related to their sensitivity to temporal synchrony, we believe that these findings support the need to further investigate the hypothesis that potential differences in temporal synchrony detection occur in EL-infants and affect later language development. However, the evidence reviewed remains inconclusive and mixed, preventing robust conclusions about this prediction.

First, not all of the reviewed studies systematically compared groups based on both autism likelihood status and diagnostic outcomes (see Elsabbagh et al., 2014 for an exception). Therefore, it remains unknown whether EL-infants, as a group, exhibit difficulties with face scanning, which would suggest that this may be a candidate mechanism shared by all infant siblings due to the increased likelihood of autism (despite potential individual differences). Or, alternatively, whether this difficulty is only present in EL-infants who go on to develop autism, suggesting that it may be an early marker of autism.

Second, future studies of infant siblings should measure selective attention to talking faces in early infancy and later outcomes, including both autism diagnosis and language development. Only one study (Chawarska et al., 2022) measured both language and autism diagnostic outcomes, and none tested associations between selective attention to talking faces, language outcomes, and autistic trait severity within the same study. This would clarify the functional specificity of difficulties in selective attention to talking faces in infant siblings.

Third, with the exception of the preliminary study by Lozano et al. (2024), there has been no longitudinal research on the patterns of change over time in selective attention to audiovisual speaking faces in EL-infants over the first year, a relevant period for the emergence of this ability in neurotypical infants. Most studies have assessed this ability at a single time-point, rather than tracking it over time. This limits our understanding of infant siblings' access to relevant cues, which is essential for learning from audiovisual redundancy and, thus, discriminating temporal synchrony.

## 5. General discussion

### 5.1. Gaps in the literature

The studies reviewed (see Table 1 for a summary) have shown that EL-infants appear to exhibit reduced sensitivity to temporal synchrony in *audiovisual speech* events (Suri et al., 2023), but this has not always been replicated (e.g., Lozano et al., 2024). When found, the exact onset of this difficulty is unclear, and greater sensitivity to this cue predicts better later productive language outcomes in this group, in contrast to neurotypical infants who showed no such relation (Suri et al., 2023). EL-infants also appear to show reduced audiovisual speech integration (Guiraud et al., 2012; Riva et al., 2022), and a trajectory of selective attention to the mouth of talking faces that is comparable to that of neurotypical infants in some studies (Chawarska et al., 2022, at 12 months; Santapuram et al., 2022, at 6–18 months) but different in others (e.g., Lozano et al., 2024, at 12 months).

The literature also suggests that, unlike neurotypical infants (Lozano et al., 2022; Tsang et al., 2018), infant siblings may lack the associations between selective attention to the mouth of talking faces and language development in the second half of the first year (Chawarska et al., 2022), suggesting a potentially different functional link that may limit the language development benefits precisely in this time period. However, this pattern was not replicated when infant siblings were tested across a wider age range (Santapuram et al., 2022), suggesting that it may be temporally limited to the end of the first year and its effects observed later on as reduced language benefits at 18 months.

The link between greater sensitivity to temporal synchrony in audiovisual speech and larger productive vocabulary may also reflect

**Table 1**
Overview of the studies reviewed and the gaps in the literature.

| | Different in EL-infants? | Earliest time-point observed? (in months) |
|---|---|---|
| **Sensitivity to Temporal Synchrony in AV Speech** | Mixed findings | Unclear (wide age-range tested) |
| **Selective Attention to the Mouth in AV Speech** | Preliminary evidence for Yes, but mixed findings | 12 |
| **AV Speech Integration** | Yes | 9 |
| **Relations Between Sensitivity to Temporal Synchrony in AV Speech and Language Outcomes** | Unclear– present in EL-infants but not in TL-infants (indicates difficulty or compensatory?) | Unclear (wide age-range tested) |
| **Relations Between Selective Attention to AV Speech and Language Outcomes** | Present in TL-infants but absent in EL-TD (not tested in EL-ASC) | 12 |
| | Present in TL and EL-infants, but indirect effect modulated by joint engagement with the caregiver and infants' prelinguistic vocal complexity. | 6–18 |

*Notes:* ASC: Autism Spectrum Condition. AV = Audiovisual. EL-infants = Infants at Elevated Likelihood for ASC. EL-TD, infants with no evidence of clinically significant symptoms of ASC; EL-ASC: infants with frank symptoms and diagnosis of ASC; TL-infants = Infants With Typical Likelihood for ASC.

a difficulty, as it was observed in 4–24-month-old infant siblings but not in neurotypical peers (Suri et al., 2023). We speculate that the absence of this association in aged-matched neurotypical infants may suggest that it developed earlier in them, and may have emerged later in infant siblings. However, the wide age range in Suri et al.'s study prevents confirmation of this possibility, and no other research has investigated when this association occurs in neurotypical infants. Alternatively, the association found by Suri and collaborators may reflect that those EL-infants who show greater sensitivity to temporal synchrony may somehow compensate for potential initial difficulties in this perceptual ability, perhaps by displaying different patterns of visual attention to talking faces than TL-infants (e.g., Lozano et al., 2024), which may benefit their later vocabulary.

The reviewed evidence also suggests that the trajectory of selective attention to the eyes declines in EL-infants compared to neurotypical infants as early as 2–6 months (Jones & Klin, 2013). However, it is unclear whether this pattern occurs in infant siblings regardless of their autism diagnosis, and whether reduced attention to the eyes affects later socio-communicative and language outcomes. Furthermore, there is insufficient evidence to conclude whether it is only infant siblings who are later diagnosed with autism who show difficulties in sensitivity to temporal synchrony and a lack of association between mouth-looking and language outcomes, or whether this extends to the whole group. Thus, taken together, the evidence supports that whereas the link between sensitivity to temporal synchrony and language outcome seems to function typically in infant siblings, the link between mouth-looking and language skills seems to be absent at least at some time-points. This finding may indicate a difficulty of infant siblings in efficiently using the audiovisual cues of this facial area for language learning. However, it remains unclear whether this occurs at a group level or only in infants with a later diagnosis of autism.

### *5.2. Ways forward: Future research directions*

The inconclusive developmental picture in the literature could benefit from some future research steps, ideally within the same study.

First, longitudinal studies with infant siblings are needed that densely track both *sensitivity to temporal synchrony* in *audiovisual speech* and *selective attention to talking faces* over the first year in the same participants. Few of the prospective longitudinal studies reviewed have tracked these abilities across multiple time-points (e.g., Droucker et al., 2013; Jones & Klin, 2013; Lozano et al., 2024); most have assessed infant siblings at a single time-point or across broad age ranges (e.g., Suri et al., 2023). Because the typical developmental trajectory of changes in selective attention to audiovisual speech appears to be closely linked to changes in temporal synchrony detection in the same event, it is fundamental to track both longitudinally in infant siblings and to test whether one or both follow a different trajectory relative to their peers. Testing group comparisons by *both* diagnostic outcome (between EL-infants with autism, EL-infants without autism, and TL-infants) and autism likelihood (EL-infants vs. TL-infants) within the same longitudinal study is important to address whether potential difficulties in sensitivity to temporal synchrony in audiovisual speech, audiovisual speech integration, and selective attention to talking faces are specific to autism or rather shared by infant siblings as a whole group.

Lozano et al. (2024) conducted such a design during the first year of life (at 4, 8 and 12 months) and observed reduced mouth-looking in *audiovisual fluent speech* in EL-infants only at 12 months, but no evidence for reduced temporal synchrony detection. Together with these findings, the lack of association between mouth-looking at 12 months and language outcomes at 18 months (Chawarska et al., 2022), and the early decline in eyes-looking observed by Jones and Klin (2013) at 2–6 months, may suggest that infant siblings may show different timing in the trajectory of preference for the eyes and mouth. However, this remains unclear due to the lack of high-powered longitudinal studies that follow these abilities in the first year *and* later language and diagnostic outcomes. Thus, the preliminary findings by Lozano et al. need to be replicated in a large sample. It is also necessary to analyze within-individual associations between performance on the two skills and later language and diagnostic outcomes.

Second, there is a need to investigate longitudinal associations between temporal synchrony detection in audiovisual speech, selective attention to the eyes and mouth, and audiovisual speech integration across the first year, all in the same event (ideally, *audiovisual fluent speech,* where evidence for group differences has been found; Lozano et al., 2024). Perceptual and attentional abilities are closely intertwined in typical development, so it is of interest to investigate whether difficulties in one or more may cascade into difficulties in the others (see Todd & Bahrick, 2022 for supporting evidence in autistic children). Although temporal synchrony detection would be the primary difficulty that would cascade into difficulties in audiovisual integration (see Bahrick & Todd, 2012), no research has tested this potential developmental pathway in EL-infants.

Third, it is crucial to test the longitudinal links between these three abilities during the first year and language outcomes in toddlerhood in order to identify potential difficulties in these functional relations compared to those observed in neurotypical infants (e.g., Edgar et al., 2023). Additionally, it is important to investigate how these abilities relate to later autism outcomes and trait severity in toddlerhood (at 24–36 months). This will help to assess whether difficulties in each ability represent cumulative risk factors for language difficulties, which are also present in the general population and other neurodevelopmental conditions, or early markers of autism. Aggregated difficulties in all of these abilities may occur only in infant siblings with a diagnosis of autism in the first year, affecting their later language development. Conversely, isolated difficulties in each of these abilities may occur in infant siblings as a whole and may not be specific to autism. Unlike infants with a diagnosis of autism, those without it may exhibit milder difficulties in language development, potentially compensating for them and restoring or approaching the typical developmental trajectory through canalization (Elsabbagh & Johnson, 2010).

Canalization to ameliorate these difficulties may occur through several mechanisms. Infant siblings as a group may have a genetic susceptibility to differences in these three abilities, but for those without a diagnosis of autism, the typical developmental trajectory could be restored through compensatory brain plasticity, environmental factors, or niche construction (i.e., the process by which individuals construct an environment that fits their processing style; Johnson et al., 2015). For example, if the typical pathway of audiovisual integration in speech is compromised in infant siblings due to the perceptual complexity of this event, a more 'auditory' processing pathway (i.e., characterized by less looking at the mouth than neurotypical peers) may emerge to maintain functionality in language development, even though this may reduce sensitivity to temporal synchrony due to lack of the benefits from attending to visual speech cues. This might explain why studies have found reduced audiovisual integration (Guiraud et al., 2012; Riva et al., 2022) and reduced sensitivity to temporal synchrony in infants' siblings, while this latter difference unexpectedly predicts better language outcomes, with larger temporal binding windows associated with better vocabulary only in this group, but not in neurotypical infants (Suri et al., 2023).

Environmental factors may also drive the canalization process. For example, caregivers may adapt to this alternative developmental pathway of more 'auditory' based audiovisual speech processing by supporting joint engagement and vocal complexity in face-to-face communication (e.g., see Quigley et al., 2016 for evidence from naturalistic interactions that mothers and EL-infants maintain the typical pitch of Infant Directed Speech longer in development than mothers and TL-infants). This may explain why in some studies infant siblings show reduced mouth-looking (Lozano et al., 2024), while in others only those with more supportive joint engagement from their caregivers and greater vocal complexity show more mouth-looking and better expressive language (Santapuram et al., 2022). As suggested by Santapuram and collaborators, infants' increased mouth-looking during face-to-face interactions may help them develop language skills by showing caregivers that they are 'tuned in' to their communication and developmentally ready to interact. This may encourage caregivers to provide joint engagement, where they create scaffolded interactions that involve reciprocal engagement around shared objects or events. These interactions provide a setting in which the infants' communication becomes meaningful and relevant, making it an ideal context for facilitating early pre-linguistic vocalizations that can lead to stronger expressive language development over time.

Finally, we still need to clarify whether basic perceptual and attentional difficulties alone explain the language difficulties associated with autism, or if core difficulties in social attention play a role. While the field is moving toward accepting perceptual difficulties as a self-sufficient explanation for core social difficulties in emerging autism and subclinical language difficulties extended to infant siblings (Falck-Ytter & Bussu, 2023), this is still a controversial issue that requires further research. Addressing all of these gaps will have a significant impact on the field.

## 6. Theoretical implications

### *6.1. Towards solving the puzzle: Targeting new domain-general difficulties*

Infants later diagnosed with autism show few overt behavioral signs of difficulties that do not fit the cognitive profile associated with autistic traits (Elsabbagh & Johnson, 2010; 2016), challenging models of autism development. This is known as the 'first-year puzzle' and it has been addressed by two theories: the *social-first hypotheses* and *domain-general accounts*.

*Social-first hypotheses* propose that the social features characteristic of autism arise from different social orienting biases (Schultz, 2005), which should therefore be explored in early components of the social brain (i.e., a network of regions involved in processing social information). However, evidence suggests that there are no differences between infants with a diagnosis of autism and neurotypical peers on several measures of social orientation in the first year, such as face orienting (Elsabbagh et al., 2013b; Palomo et al., 2022), or face scanning (Elsabbagh et al., 2014). Alternatively, *domain-general accounts* suggest that the earliest precursors of social difficulties may not be restricted to the social domain (Campos, 2018; Elsabbagh & Johnson, 2016; Gliga et al., 2014; Johnson et al., 2015; Johnson, 2017). Instead, widespread difficulties in different systems would precede social features, which is supported by studies showing attentional (Elsabbagh et al., 2013a), perceptual (Piccardi et al., 2021), visual (Gliga et al., 2015) and sensorimotor (Iverson & Wozniak, 2007) difficulties in infants with a later autism of diagnosis. Domain-general accounts mainly include Karmiloff-Smith's Neuroconstructivism (although it would be more accurate to call it domain-relevant; see the underlying rationale in Karmiloff-Smith, 1998; 2015; Campos, 2018; Elsabbagh & Johnson, 2016; Gliga et al., 2014; Johnson et al., 2015), and the *Intersensory Impairment Hypothesis* for autism by Bahrick and collaborators (hereafter, *IIH*; Bahrick, 2010; Bahrick & Todd, 2012), which represents an extension of the IRH (Bahrick & Lickliter, 2002; 2014).

While *domain-general accounts* are more consistent with current evidence than the *social first hypotheses*, explaining how broad difficulties *differentially* affect the social domain is a challenge. The social world contains fast, complex, dynamic, and highly unpredictable stimuli, such as human faces or audiovisual speech. Difficulties in the ability to deal with this uncertainty may hinder infants' adaptation to the social environment, potentially reducing their social engagement. Even subtle initial difficulties may result in '*the early environment being sampled with poor fidelity, with a particular cost to the most dynamic and least easily predictable elements of the external environment*' (Johnson et al., 2015, p. 434). While 'widespread difficulties' may affect social and non-social stimuli, social stimuli would be differentially more affected due to their complexity. This mismatch between infants' abilities and the processing demands of their social environment could lead them to prioritize simpler, non-social stimuli over complex social ones, at the cost of reduced attunement to the social world (Gliga et al., 2014; Johnson, 2017). This could cascade into the social features observed in autism and the subclinical language difficulties reported in infants' siblings as a whole (Belteki et al., 2022).

The *IIH* hypothesizes that reduced detection of temporal asynchrony would cascade into difficulties with audiovisual integration and, ultimately, into the social-communicative features that are at the core of autism. It may also cascade into subclinical developmental trajectories of language difficulties associated with autism (Bahrick, 2010; Bahrick & Todd, 2012). An early difficulty in detecting temporal synchrony could affect how redundancy guides infants' selective attention to relevant socio-communicative events, losing the benefits of unified perception and gains in this social domain (Bahrick, 2010).

This hypothesis is unique compared to other domain-general accounts in guiding early autism research, because (1) it makes a very specific prediction about which difficulties in very specific *domain-general mechanisms*—primarily temporal synchrony detection, but also selective attention—may lead to *domain-specific difficulties* associated with autism (in socio-communicative skills, including language), (2) it provides a *developmental approach* by linking subtle early difficulties in a low-level perceptual and attentional mechanism and their impact on later higher-order skills associated with autism, and (3), by highlighting the complex perceptual demands of social contexts on infants, it explains *why* temporal synchrony detection is *differentially* fundamental for infants to succeed in social contexts (vs. non-social), and, more importantly, *why* a difficulty in this perceptual mechanism would *differentially* affect social (vs. non-social) skills in autistic individuals.

As we have shown in this review, taken together, the evidence in infant siblings only partially supports the *IIH.* Contrary to predictions, the link between sensitivity to temporal synchrony and language outcomes appears to be typically functioning in infant siblings. However, as predicted, the link between looking at the mouth and language skills appears to be lacking—at least at some timepoints—suggesting a potential difficulty in efficiently using the audiovisual cues of this facial region for language learning. However, it remains unclear whether this occurs at a group level or only in infants who are later diagnosed with autism.

Evidence for the *IIH* is also mixed when more broadly looking at the literature on autistic individuals across the lifespan. Although evidence is mixed (as age and stimulus type modulate this difficulty; Feldman et al., 2018), in support of this hypotheses, previous research has found differential sensitivity to temporal synchrony across the lifespan (Stevenson et al., 2016), with reduced temporal asynchrony observed in autistic children aged 3–4 years (Irwin et al., 2011; Mongillo et al., 2008). It starts early and continues through childhood and adolescence, but not into adulthood (de Boer-Schellekens et al., 2013a), supporting the *IIH* that sensitivity to temporal synchrony may be affected in autism, although it is not a core feature.

Although this difficulty also occurs with non-speech stimuli (e.g., flashes and beeps; Beker et al., 2018; Foss-Feig et al., 2010; but see de Boer-Schellekens et al., 2013a,b for mixed evidence), it is more pronounced with *audiovisual speech* (de Boer-Schellekens et al., 2013b; Stevenson et al., 2014; 2018). Some studies have found temporal synchrony detection differently functions exclusively to audiovisual talking faces and not to objects impacting a surface. For example, Bebko et al. (2006) reported differential responses to temporal synchrony only for linguistic stimuli (vs. objects) in 4-to 6-year-old autistic children. Similarly, Righi et al. (2018) found reduced sensitivity to multiple temporal asynchronies in talking faces uttering fluent speech among 2-to-9-year-olds. Todd and Bahrick (2022) replicated this difficulty at ages 2–5 years, demonstrating its exclusivity to audiovisual talking faces and not to objects impacting a surface. These difficulties persist into adolescence (Zhou et al., 2022) and adulthood (Beker et al., 2018), although they appear to improve over time.

Also, in support for the *IIH* (Feng et al., 2021; Stevenson et al., 2018), autistic individuals show a larger temporal binding window across different audiovisual events and ages, perceiving auditory and visual stimuli as coming from the same event over longer time intervals than neurotypicals (Foss-Feig et al., 2010; Wallace & Stevenson, 2014). Although not exclusive, this is more pronounced for *audiovisual speech* events (Zhou et al., 2022). Differences between autistic and neurotypical individuals decrease with age, becoming equivalent in adulthood for both non-social (Ainsworth & Bertone, 2023; Weiland et al., 2023) and social events (Poole et al., 2023).

However, contrary to the *IIH,* difficulties in temporal synchrony detection in autistic individuals do not consistently correlate with socio-communicative or language difficulties, raising uncertainty as to whether the former accounts for the latter. In autistic children and adolescents (aged 8–17 years), reduced perception of unity in audiovisual speech stimuli positively correlates with socio-communicative and sensory difficulties (Mongillo et al., 2008; Woynaroski et al., 2013). Similarly, reduced audiovisual speech integration is associated with the severity of autistic traits (Feldman et al., 2018), but exceptions exist, particularly in links to language difficulties (e.g., Todd & Bahrick, 2022; Weiland et al., 2023). For example, in autistic children (aged 2–9 years), the detection of temporal asynchronies in audiovisual talking faces is positively correlated with concurrent language skills, as in neurotypicals (Righi et al., 2018). Consistently, autistic children (aged 2–5 years) show reduced temporal synchrony detection in audiovisual talking faces compared to neurotypicals, but this ability predicts language outcomes in both groups (Todd & Bahrick, 2022). However, whereas in neurotypical children attentional maintenance and disengagement of attention to talking faces mediated the positive link between temporal synchrony detection and language outcomes, in autistic children, temporal synchrony did not mediate this relation. Instead, maintaining attention and detecting temporal synchrony in talking faces were unrelated and independently predicted better language outcomes. This suggests a potential disconnect between these two skills in autistic children, yet improvements in each positively influenced language development. In addition, temporal synchrony detection in talking faces did not predict the severity of autistic traits.

Taken together, the evidence from infant siblings and autistic individuals only partially supports the *IIH*, leaving unresolved issues in understanding sensitivity to temporal synchrony in autism. First, the developmental patterns leading to difficulties in audiovisual processing in autism remain unexplored; they are more pronounced at younger ages and decline in adolescence as autistic individuals 'catch up' with their peers (Taylor et al., 2010), but their onset and potential changes during early development remain unknown. Second, and crucially, the relation between audiovisual processing and temporal synchrony detection across development in both autistic, infant siblings, and neurotypical populations requires further investigation. Researchers increasingly agree that, at least for *audiovisual speech*, different temporal synchrony detection in autistic individuals likely precedes and explains the difficulties in audiovisual integration for the same events (Stevenson et al., 2016; Wallace et al., 2019), but further replication is needed.

### *6.2. Domain-general difficulties may become domain-specific over time*

Despite only partial support for the *IIH*, it could still guide future research on the early developmental course of temporal synchrony difficulties in autism and their potential cascading effects, addressing some of the literature gaps identified above. Difficulties in early sensitivity to temporal synchrony and selective attention remain outside most of the *domain-general* models, with the exception of the

*IIH,* where they are central. One reason for this may be the ongoing theoretical debate about whether the difficulties found in infant siblings in some studies (e.g., Suri et al., 2023) are initially domain-general or domain-specific, and whether this specificity changes during the first year.

In this review, we have focused on discussing studies of temporal synchrony detection in *audiovisual speech* events, which is fundamental for early language acquisition, but difficulties in this ability in infant siblings with a diagnosis of autism also occur for other social events by the end of the first year (biological motion; (Falck-Ytter et al., 2018). However, most studies lack a full developmental trajectory and often measure only one time-point. For example, Suri et al. (2023) showed cross-sectionally that 4–24 months-old infant siblings exhibit difficulties in this ability for audiovisual speech events, but not for non-social events (e.g., a bouncing ball). However, this may not hold when measured at earlier and more specific time-points, rather than across a range of ages.

We propose that potential difficulties in temporal synchrony detection and selective attention may start as *domain-general,* but have

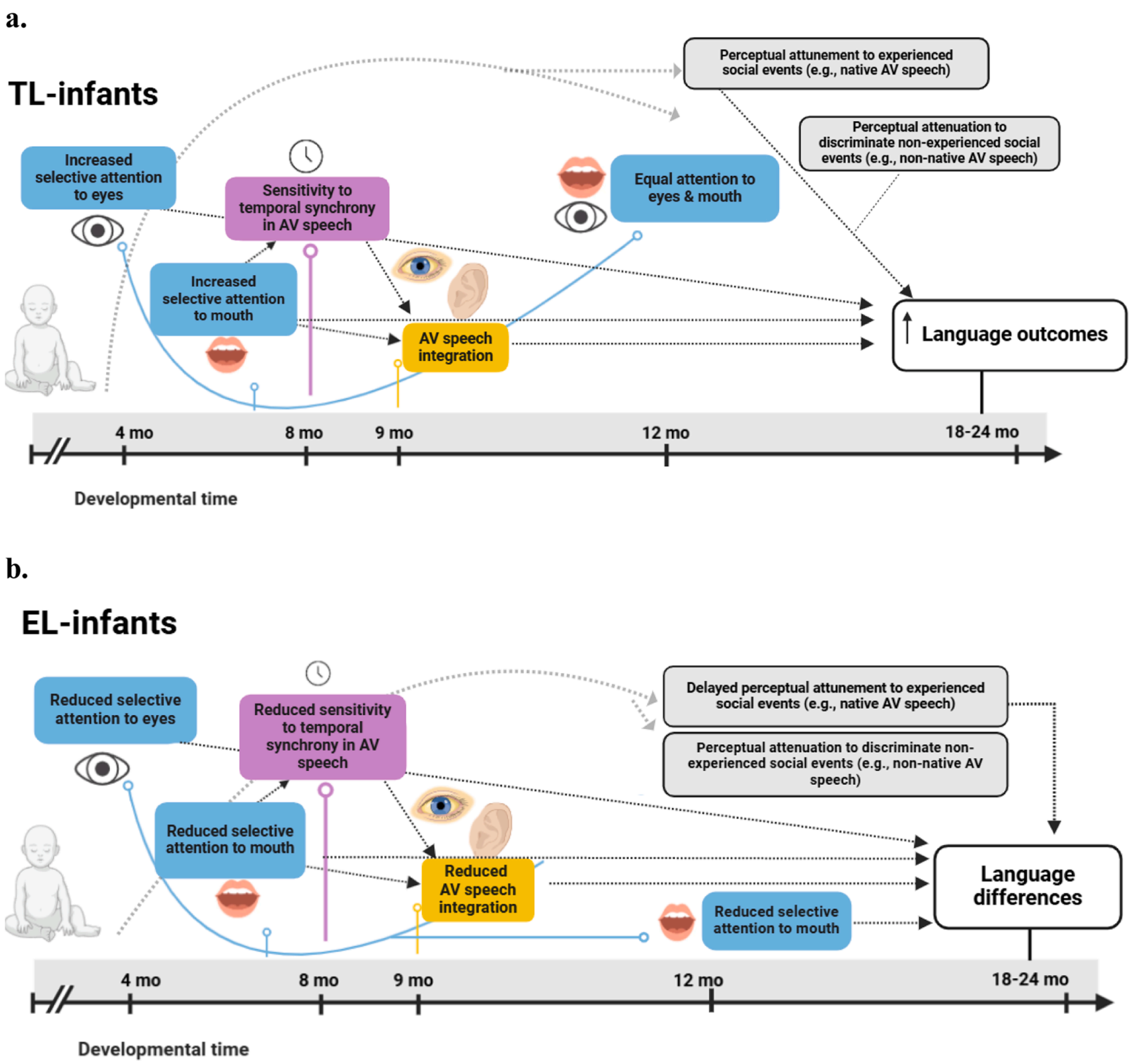


**Fig. 1.** Overview of our conceptual model illustrating developmental pathways linking sensitivity to temporal synchrony (STS, purple) and selective attention (SA, blue) in audiovisual (AV) speech to later language outcomes in TL-infants (a) and EL-infants (b). Dashed lines indicate relations across the lifespan, direct or indirect (non-causal). In EL-infants (at the group level), reduced STS and SA may directly impact language outcomes or do so via reduced AV speech integration (yellow). Although SA and STS are domain-general mechanisms, their disruption may especially delay perceptual narrowing to relevant social events like AV speech (depicted here), faces, or phonemes, as they are more complex than non-social events. Ultimately, they may cascade into language differences in EL infants, as perceptual narrowing to social events is crucial for typical language acquisition. Differences in SA and STS can also lead to different dyadic interactions, omitted here for clarity. TL: Typical Likelihood of autism; EL: Elevated Likelihood of autism; AV: Audiovisual; Mo: months. Created with BioRender.com.

greater cascading effects on infant siblings' attunement to their social environment due to the complex nature of these events, such as *audiovisual speech* (see Fig. 1 for an overview of our proposal). Indeed, autistic children show more pronounced difficulties in detecting temporal asynchronies in social versus non-social events (Falck-Ytter et al., 2013; Klin et al., 2009). However, it remains unclear whether this asymmetry extends to infant siblings as a whole and, if so, *when* it emerges (possibly between 4 and 24 months, as suggested by Suri et al., 2023). Again, we suggest that a longitudinal approach is needed to track whether domain-general difficulties *become* domain-specific over time (Elsabbagh & Johnson, 2016).

Importantly, with the exception of Suri et al. (2023), to our knowledge, no other study has compared temporal synchrony detection between social and non-social events in infant siblings. Systematically investigating early difficulties in *both* sensitivity to temporal synchrony and selective attention in EL-infants in audiovisual speech and other social events compared to non-social events during the first year might also help to solve the *first-year puzzle* for two reasons.

First, as in neurotypical infants (Edgar et al., 2023), sensitivity to temporal synchrony may be differentially important for EL-infants to engage effectively in social contexts. Difficulties in this mechanism could disproportionately affect the processing of social events relative to non-social events. Consistent with *domain-general* accounts, this differential effect may be due to the fact that social events are complex, multimodal, unpredictable, and highly temporally variable. Reduced sensitivity to temporal synchrony, which is essential for linking auditory and visual stimuli, could make these events *'confusing, aversive, and well beyond the optimal range of sensory stimulation'* (Bahrick & Todd, 2012, p. 662) rather than perceptually unified, potentially affecting infants' social engagement. The same reasoning can be extended to audiovisual speech events. Given its complexity, if EL-infants develop difficulties in detecting temporal synchrony early in life, the demands of processing this event may exceed their optimal range of sensory stimulation.

One possibility is that EL-infants may follow an alternative route of processing audiovisual speech as an *'adaptation'* to simplify this complex event (see Johnson et al., 2015). We hypothesize that EL-infants may avoid relying on visual cues from the articulating mouth, as these may not facilitate the processing of redundant information (see Zhao et al., 2023 for evidence in autistic children). Instead, EL-infants may use a more 'auditory-based' form of audiovisual speech processing. While this alternative path may be better suited to their perceptual abilities, it may limit their access to important visual linguistic cues, thereby constraining language learning. This is supported by the group-level language difficulties exhibited by EL-infants (Belteki et al., 2022), the lack of associations between mouth-looking at the end of the first year and language outcomes in toddlerhood (Chawarska et al., 2022), and the reduced sensitivity to temporal synchrony in EL-infants only for *audiovisual speech* events, not objects (Suri et al., 2023). However, the current designs and ages tested have only provided evidence for group differences in mouth-looking in a preliminary longitudinal study (Lozano et al., 2024). To further test this possibility, high-powered infant sibling studies are needed to track this ability and temporal synchrony detection in social and non-social events in the first year, and language and diagnostic outcomes in toddlerhood (see Todd & Bahrick, 2022 for a cross-sectional example in autistic children).

Second, selective attention is a domain-general skill that is critical for infants to sample their social contexts (Bahrick, 2010). Like sensitivity to temporal synchrony, this mechanism helps infants handle a dynamic multimodal environment, which supports language acquisition (Edgar et al., 2023; Tsang et al., 2018; Young et al., 2009). Investigating longitudinally whether selective attention to *both* social (e.g., audiovisual speech) and non-social (e.g., objects) events becomes altered in infant siblings during the first year and, if so, its implications for later language and autism outcomes will shed light on the *first-year puzzle*. Designing selective attention tasks that use non-social stimuli that are comparable to social stimuli in complexity, but not in content, is a key challenge that remains to be solved. Naturalistic social and non-social events differ in complexity, making it difficult to define clear criteria for equating their properties and ensuring comparable complexity.

### *6.3. Mind to focus on development itself*

Although the *IIH* has contributed to target two very specific perceptual and attentional mechanisms that may function differently in infant siblings, some important theoretical gaps remain in this approach. However, these can be addressed by other current domain-general accounts, primarily Neuroconstructivism (Campos, 2018; Johnson et al., 2015; Karmiloff-Smith, 1998).

First, it is fundamental to explain the *developmental processes* linking primary (e.g., perceptual) and derivative (e.g., linguistic) difficulties in EL-autism infants. In contrast to the *IIH*, Neuroconstructivism suggests *specialization* as the underlying affected *developmental mechanism* and *perceptual narrowing* as the fundamental affected *process*, but this has not been tested.

Second, it is necessary to formulate accurate theoretical predictions in terms of developmental *time* (i.e., *when* should we expect to see difficulties in these two mechanisms in EL-infants?). We propose that the Neuroconstructivist approach suggests relevant time-points that need to be explored. This account hypothesizes that EL-infants would exhibit group-level delays in the perceptual narrowing of relevant social events in their environment, including audiovisual speech, native language phonemes, or the same-race faces of people with whom they interact (Campos et al., 2019; Johnson et al., 2015). As neurotypical infants complete this process by the end of the first year (Maurer & Werker, 2014), future research should focus on investigating the performance of EL-infants at this time-point and compare it with their performance at the classical pre-perceptual narrowing time-points (4 and 8 months)—for example, as in Lozano et al. (2024), although these authors did not test perceptual narrowing in audiovisual speech. Furthermore, longitudinal studies with infant siblings should investigate whether potential difficulties in temporal synchrony detection and selective attention in *audiovisual speech* delay infants' perceptual narrowing in this event.

Third, it is not so clear that the comparison needed to examine domain-general vs. domain-specific difficulties is necessarily social vs. non-social events, as assumed by the *IIH*. Neuroconstructivism suggests the need to rethink comparing infant siblings' performance on social vs. non-social stimuli, advocating instead comparisons between experienced vs. non-experienced social events (i.e., native vs. non-native, including audiovisual speech; Campos, 2018).

Fourth, a more comprehensive developmental picture of the cascading effects of perceptual and attentional difficulties in infant siblings is needed. From a Neuroconstructivist perspective, the initial constraints in infant siblings may occur internally in their neurocognitive system (e.g., difficulties in temporal synchrony detection), but also externally (e.g., the amount of audiovisual speech experience received from caregivers) and in infant-caregiver interactions (internal-external). Therefore, future studies of temporal synchrony detection and selective attention should ideally track variables at these three levels, rather than focusing primarily on the internal level. For example, at the external level, we could manipulate audiovisual redundancy in infant-caregiver dyads by presenting infants with a silent 'talking' caregiver and comparing their preferences for eyes and mouth with those in a condition where the caregiver provides audiovisual speech. At the internal-external level, studying infants of blind caregivers, who do not experience the visual cues that typically accompany auditory speech in a talking face—such as eye gaze, eyebrow movements, or gestures—provides an interesting opportunity to explore, in a natural setting, the differential effects of audiovisual speech features versus dyadic features on infants' attention to the eyes and mouth. Furthermore, it is important to investigate whether reciprocity and turn-taking—dimensions related to temporal synchrony at the dyadic level—influence infants' preferences and engagement with talking faces, which may be partially influenced by the caregiver's behavior (e.g., mutual gaze, facial expressions, body movements).

Finally, we need to address the potential non-specificity of difficulties in temporal synchrony detection and selective attention in autistic individuals and infant siblings. While the *IIH* focuses the predictions on autism, children with Specific Language Impairment, preterm infants, or infants with Fragile X also exhibit them (D'Souza et al., 2016; Pons et al., 2013). From Neuroconstructivism and cumulative risk factor models of emerging autism (Elsabbagh, 2020), a complete picture requires the investigation of potential common and divergent trajectories of these two mechanisms across neurodevelopmental conditions, as well as compensatory pathways. Therefore, it will be important to conduct cross-syndrome studies that longitudinally track these two abilities throughout the first year of life.

### 6.4. Broadening theoretical frameworks

Future studies should expand their theoretical frameworks beyond domain-general accounts of autism to interpret the findings. Bayesian models of perception in autism may provide insights into the inconclusive results reviewed (e.g., Pellicano & Burr, 2012). According to these models, the perception of individuals with autism is less influenced by their prior beliefs about the environment, their so-called 'priors' (see Chrysaitis & Seriès, 2023, for a systematic review). Could the detection of temporal synchrony and audiovisual speech integration be considered as 'priors' of a different nature? This perspective could explain why, at the group level, infant siblings show reduced susceptibility to the McGurk illusion for audiovisual speech syllables at 9 and 12 months (Guiraud et al., 2012; Riva et al., 2022), whereas some studies show difficulties in temporal synchrony detection for the same event between 4 and 24 months (Suri et al., 2023), thus starting slightly earlier. Could it be that infant siblings show reduced *learning* of 'priors'? Or do they exhibit difficulties with reduced *adaptation*, defined as "the dynamic process by which neural sensitivity is continuously recalibrated to 'match' the characteristics of the current environment" (Pellicano & Burr, 2012, p. 504)?

Interestingly, social priors are more affected than non-social priors in individuals with autism, consistent with the domain asymmetry we predict in infant siblings after the first year. Alternative perceptual models may explain why reduced temporal synchrony detection in this population differs in timing from audiovisual integration difficulties (Guiraud et al., 2012; Suri et al., 2023). From the ecological approach by Gibson and Pick (2000), it could be argued that infant siblings may show an affected 'differentiation process', whereby neurotypical infants learn from global perceptual cues (i.e., amodal or unspecific from any sensory modality) to increasingly more specific perceptual cues.

Sensory explanations should also be considered. One unexplored possibility is that sensory processing may influence selective attention in infant siblings, particularly with regard to social stimuli (see Thye et al., 2018 for evidence in autistic individuals). Increased sensory responsiveness could make certain stimuli (particularly social stimuli such as audiovisual speech) more salient and attentionally demanding. Those EL-infants who show hyper-responsiveness/hypo-responsiveness to specific stimuli may find it difficult to shift their attention away from these stimuli (e.g., eyes versus the mouth) and redirect their attention to alternative (non-social) stimuli. Differential sensory responses may also affect audiovisual integration/temporal synchrony detection (as in Riva et al., 2022). Thus, EL-infants may perceive a stimulus without efficiently attending to it, making it difficult to determine whether a given gaze response is driven by sensory or attentional difficulties. Future studies should explore the potential pathways linking EL-infants' sensory responsiveness, selective attention, and audiovisual speech processing, and the downstream effects on socio-communicative skills, including language skills (see Feldman et al., 2024, for partial supporting evidence). For example, we could investigate whether sensory subgroups associated with greater later socio-communicative difficulties and restricted/repetitive behaviors in early childhood, such as high sensitivity (see Riboldi et al., 2024), are associated with earlier difficulties in detecting temporal synchrony in audiovisual speech, audiovisual integration, and selective attention to the mouth in infancy.

Furthermore, as touch is often synchronized with audiovisual speech during dyadic interactions, future studies could investigate whether adding touch to audiovisual speech enhances its processing and temporal synchrony detection, and whether this differs in infants' siblings compared to neurotypical peers. For example, we could study whether exposing infants to audio-haptic speech—-where they touch the caregiver's lips while listening to speech (see, e.g., Treille et al., 2014 for evidence in adults)—enhances their audiovisual integration in the McGurk effect paradigm. Naturalistic observations might be a valuable methodological approach to exploring the importance of touch in audiovisual speech procesisng during interactions. In addition, we could investigate whether infants' articulatory configurations (i.e., oral-motor movements) affect their audiovisual speech integration, as has already been shown to affect their auditory speech discrimination (Bruderer et al., 2015).

Finally, the field should consider sensory explanations at the neural level for a more integrative approach. EL-infants may show difficulties coordinating cortical connectivity in response to audiovisual speech, affecting the mapping of sensory input and higher-level skills (see Jochaut et al., 2015 for evidence in adults with autism). Future research should also investigate the potential differential associations between cortical connectivity, tracking accuracy of audiovisual speech (i.e., neural tracking of the speech envelope), and gaze to the mouth in infant siblings. Alternatively, but compatibly, an imbalance of excitation and inhibition in the audiovisual cortex (via a reduction of GABAergic inhibition) could affect the temporal binding window of audiovisual perception in infant siblings, as shown in animal models (Schormans & Allman, 2023), but its potential generalization to audiovisual speech events and human infants remains unexplored.

## 7. Conclusions

Despite strong theoretical reasons for the investigation of sensitivity to temporal synchrony in audiovisual speech in EL-infants, direct and indirect evidence remains limited. The research reviewed suggests evidence for reduced sensitivity to temporal synchrony in audiovisual speech in infant siblings, but this is not found systematically. The functional link between this perceptual ability and language outcomes does not appear to be affected, suggesting that it may not be a candidate mechanism to explain group-level language difficulties. However, a potential lack of relation between the trajectory of selective attention to the articulating mouth and later language outcomes may be a plausible candidate for language difficulties, but further investigation is needed to determine whether this is specific to infant siblings with a diagnosis of autism. Longitudinal eye-tracking studies following the developmental trajectories of both mechanisms across domains in EL-infants during the first year, and both language and autism-related outcomes in toddlerhood, are important future steps needed.

## Author contributions

This article was part of the doctoral thesis carried out by IL at Universidad Autónoma de Madrid with the academic supervision of RC and MB. The three authors are responsible for both the content, organization, and final writing of this manuscript. **CRediT authorship contribution statement: Itziar Lozano:** Conceptualization, Funding acquisition, Writing - original draft, Writing - review & editing. **Mercedes Belinchón:** Conceptualization, Supervision, Writing – review & editing. **Ruth Campos:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing - review & editing.

## Funding

This work was funded by the Spanish Ministry of Education, Culture and Sports [grant FPU13/03508 grant FPU 2016 for short-term international stays] and Universidad Autónoma de Madrid [UAM 2018 mobility subprogram], awarded to IL. The work was also supported by funds from the Spanish Ministry of Economy and Competitiveness and the European Regional Developmental Fund (ERDF) [TRABERITEA Project – grant PSI2015–66509-P MINECO/FEDER] and the Ministry of Science, Innovation, and State Agency of Research [grant PID2020–117087GB-I00], awarded to RC. Additional work of this research has been funded by a National Science Centre of Poland grant Sonatina 7 [Grant number 2023/48/C/HS6/00264], awarded to IL. Any views expressed are those of the author (s) and not necessarily those of the funders. The funders had no role in study design, data collection and analysis, the decision to publish or manuscript preparation.

## CRediT authorship contribution statement

**Itziar Lozano:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization. **Mercedes Belinchón:** Writing – review & editing, Supervision, Conceptualization. **Ruth Campos:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

## Declaration of Competing Interest

None.

## Acknowledgments

Itziar Lozano would like to thank Chris Winspear for his help with the English proofreading and Prof. Lorraine E. Bahrick for the insightful theoretical discussions in her Infant Development Laboratory.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.infbeh.2024.102026.

## Data availability

No data was used for the research described in the article.

## References


Ainsworth, K., & Bertone, A. (2023). Audiovisual temporal binding window narrows with age in autistic individuals. *Autism Research, 16*(2), 355–363. https://doi.org/10.1002/aur.2860

American Psychiatric Association. (2022). *Diagnostic and Statistical manual of mental disorder* (5th ed.—Text revision). Arlington, VA: American Psychiatric Publishing.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development, 6*(4), 429–451. https://doi.org/10.1016/S0163-6383(83)90241-2

Bahrick, L. E. (2010). Intermodal perception and selective attention to intersensory redundancy: Implications for typical social developmental and autism. In J. G. Bremner, & T. D. Wachs (Eds.), *The Wiley- Blackwell handbook of infant development* (2nd ed., pp. 120–166). Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444327564.ch4.

Bahrick, L., & Hollich, G. (2008). Intermodal perception. In M. Haith y, & J. Benson (Eds.), *Encyclopedia of Infant and Early Childhood Development*. San Diego, California USA: Academic Press. https://doi.org/10.1016/B978-012370877-9.00086-4.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*(2), 190. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. *Advances in Child Development and Behavior, 30*, 153–189. https://doi.org/10.1016/S0065-2407(02)80041-6

Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current Directions in Psychological Science, 23*(6), 414–420. https://doi.org/10.1177/0963721414549187

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 In. B. E. Stein (Ed.), *The new handbook of multisensory processes* (pp. 657–674). Cambridge, MA: MIT Press.

Bai, D., Yip, B. H. K., Windham, G. C., Sourander, A., Francis, R., Yoffe, R., … Gissler, M. (2019). Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA Psychiatry, 76*(10), 1035–1043. https://doi.org/10.1001/jamapsychiatry.2019.1411

Bastianello, T., Keren-Portnoy, T., Majorano, M., & Vihman, M. (2022). Infant looking preferences towards dynamic faces: A systematic review. *Infant Behavior and Development, 67*, Article 101709. https://doi.org/10.1016/j.infbeh.2022.101709

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

Beker, S., Foxe, J. J., & Molholm, S. (2018). Ripe for solution: Delayed development of multisensory processing in autism and its remediation. *Neuroscience and Biobehavioral Reviews, 84*, 182–192. https://doi.org/10.1016/j.neubiorev.2017.11.008

Belteki, Z., Lumbreras, R., Fico, K., Haman, E., & Junge, C. (2022). The vocabulary of infants with an elevated likelihood and diagnosis of autism spectrum disorder: A systematic review and meta-analysis of infant language studies using the CDI and MSEL. *International Journal of Environmental Research and Public Health, 19*(3), 1469. https://doi.org/10.3390/ijerph19031469

Bhat, A. N., Landa, R. J., & Galloway, J. C. (2011). Current perspectives on motor functioning in infants, children, and adults with autism spectrum disorders. *Physical Therapy, 91*(7), 1116–1129. https://doi.org/10.2522/ptj.20100294

Birulés, J., Goupil, L., Josse, J., & Fort, M. (2023). The role of talking faces in infant language learning: Mind the gap between screen-based settings and real-life communicative interactions. *Brain Sciences, 13*(8), 1167. https://doi.org/10.3390/brainsci13081167

Blasi, A., Lloyd-Fox, S., Elwell, C., Charman, T., Murphy, D. G. M., & Johnson, M. H. (2013). Reduced neural sensitivity to social stimuli in infants at risk for autism. *Proceedings of the Royal Society B: Biological Sciences, 280*(1758). https://doi.org/10.1098/rspb.2012.3026

de Boer-Schellekens, L., Eussen, M., & Vroomen, J. (2013b). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience, 7*(8). https://doi.org/10.3389/fnint.2013.00008

de Boer-Schellekens, L., Keetels, M., Eussen, M., & Vroomen, J. (2013a). No evidence for impaired multisensory integration of low-level audiovisual stimuli in adolescents and young adults with autism spectrum disorders. *Neuropsychologia, 51*(14), 3004–3013. https://doi.org/10.1016/j.neuropsychologia.2013.10.005

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 112*(44), 13531–13536. https://doi.org/10.1073/pnas.1508631112

Campos, R. (2018). If you want to get ahead, get a good master. Annette Karmiloff-Smith: the developmental perspective/Si quieres avanzar, ten una buena maestra. Annette Karmiloff-Smith: la mirada desde el desarrollo. *Infancia York Aprendizaje, 41*(1), 90–137. https://doi.org/10.1080/02103702.2017.1401318

Campos, R., Nieto, C., & Núñez, M. (2019). Research domain criteria from neuroconstructivism: A developmental view on mental disorders. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(3), Article e1491. https://doi.org/10.1002/wcs.1491

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

Charman, T., Young, G. S., Brian, J., Carter, A., Carver, L. J., Chawarska, K., … Hertz-Picciotto, I. (2017). Non-ASD outcomes at 36 months in siblings at familial risk for autism spectrum disorder (ASD): A baby siblings research consortium (BSRC) study. *Autism Research, 10*(1), 169–178. https://doi.org/10.1002/aur.1669

Chawarska, K., Lewkowicz, D., Feiner, H., Macari, S., & Vernetti, A. (2022). Attention to audiovisual speech does not facilitate language acquisition in infants with familial history of autism. *Journal of Child Psychology and Psychiatry, 63*(12), 1466–1476. https://doi.org/10.1111/jcpp.13595

Chawarska, K., Macari, S., Powell, K., DiNicola, L., & Shic, F. (2016). Enhanced social attention in female infant siblings at risk for autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 55*(3), 188–195. https://doi.org/10.1016/j.jaac.2015.11.016

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry, 74*(3), 195–203. https://doi.org/10.1016/j.biopsych.2012.11.022

Chrysaitis, N. A., & Seriès, P. (2023). 10 years of Bayesian theories of autism: A comprehensive review. *Neuroscience and Biobehavioral Reviews*. , Article 105022. https://doi.org/10.1016/j.neubiorev.2022.105022

Cox, C. M. M., Keren-Portnoy, T., Roepstorff, A., & Fusaroli, R. (2022). A Bayesian meta-analysis of infants' ability to perceive audio–visual congruence for speech. *Infancy, 27*(1), 67–96. https://doi.org/10.1111/infa.124

D'Souza, D., D'Souza, H., Johnson, M. H., & Karmiloff-Smith, A. (2016). Audio-visual speech perception in infants and toddlers with Down syndrome, fragile X syndrome, and Williams syndrome. *Infant Behavior and Development, 44*, 249–262. https://doi.org/10.1016/j.infbeh.2016.07.002

Droucker, D., Curtin, S., & Vouloumanos, A. (2013). Linking infant-directed speech and face preferences to language outcomes in infants at risk for autism spectrum disorder. *Journal of Speech, Language, and Hearing Research, 56*(2), 567–576. https://doi.org/10.1044/1092-4388(2012/11-0266

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2022). Intersensory matching of faces and voices in infancy predicts language outcomes in young children. *Developmental Psychology, 58*(8), 1413. https://doi.org/10.1037/dev0001375

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2023). Intersensory processing of social events in 6-month-olds predicts language outcomes at 18, 24, and 36 months of age. *Infancy, 28*(3), 569–596. https://doi.org/10.1111/infa.12533

Eigsti, I. M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders, 5*(2), 681–691. https://doi.org/10.1016/j.rasd.2010.09.001

Elsabbagh, M. (2020). Linking risk factors and outcomes in autism spectrum disorder: Is there evidence for resilience? *BMJ, 368*, l6880. https://doi.org/10.1136/bmj.l6880

Elsabbagh, M., Bedford, R., Senju, A., Charman, T., Pickles, A., & Johnson, M. H. (2014). What you see is what you get: Contextual modulation of face scanning in typical and atypical development. *Social Cognitive and Affective Neuroscience, 9*(4), 538–543. https://doi.org/10.1093/scan/nst012

Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., Johnson, M. H., & BASIS Team. (2013a). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry, 74*(3), 189–194. https://doi.org/10.1016/j.biopsych.2012.11.030

Elsabbagh, M., Gliga, T., Pickles, A., Hudry, K., Charman, T., Johnson, M. H., & BASIS Team. (2013b). The development of face orienting mechanisms in infants at-risk for autism. *Behavioral Brain Research, 251*, 147–154. https://doi.org/10.1016/j.bbr.2012.07.030

Elsabbagh, M., & Johnson, M. H. (2010). Getting answers from babies about autism. *Trends in Cognitive Science, 14*(2), 81–87. https://doi.org/10.1016/j.tics.2009.12.005

Elsabbagh, M., & Johnson, M. H. (2016). Autism and the social brain: The first-year puzzle. *Biological Psychiatry, 80*(2), 94–99. https://doi.org/10.1016/j.biopsych.2016.02.019

Elsabbagh, M., Volein, A., Csibra, G., Holmboe, K., Garwood, H., Tucker, L., … Baird, G. (2009b). Neural correlates of eye gaze processing in the infant broader autism phenotype. *Biological Psychiatry, 65*(1), 31–38. https://doi.org/10.1016/j.biopsych.2008.09.034

Elsabbagh, M., Volein, A., Holmboe, K., Tucker, L., Csibra, G., Baron-Cohen, S., … Johnson, M. H. (2009a). Visual orienting in the early broader autism phenotype: Disengagement and facilitation. *Journal of Child Psychology and Psychiatry, 50*(5), 637–642. https://doi.org/10.1111/j.1469-7610.2008.02051.x

Falck-Ytter, T., & Bussu, G. (2023). The sensory-first account of autism. *Neuroscience and Biobehavioral Reviews, 153*, Article 105405. https://doi.org/10.1016/j.neubiorev.2023.105405

Falck-Ytter, T., Kleberg, J. L., Portugal, A. M., & Thorup, E. (2022). Social attention: Developmental foundations and relevance for autism spectrum disorder. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2022.09.035

Falck-Ytter, T., Rehnberg, E., & Bölte, S. (2013). Lack of visual orienting to biological motion and audiovisual synchrony in 3-year-olds with autism. *PLOS One, 8*(7), Article e68816. https://doi.org/10.1371/journal.pone.0068816

Falck-Ytter, T., Nyström, P., Gredebäck, G., Gliga, T., Bölte, S., EASE team, … Thorup, E. (2018). Reduced orienting to audiovisual synchrony in infancy predicts autism diagnosis at 3 years of age. *Journal of Child Psychology and Psychiatry, 59*(8), 872–880. https://doi.org/10.1111/jcpp.12863

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 and Biobehavioral Reviews, 95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Feldman, J. I., Garla, V., Dunham, K., et al. (2024). Longitudinal relations between early sensory responsiveness and later communication in infants with autistic and non-autistic siblings. *Journal of Autism and Developmental Disorders, 54*(2), 594–606. https://doi.org/10.1007/s10803-022-05817-3

Feng, S., Lu, H., Fang, J., Li, X., Yi, L., & Chen, L. (2021). Audiovisual speech perception and its relation with temporal processing in children with and without autism. *Reading and Writing*, 1–22. https://doi.org/10.1007/s11145-021-10200-2

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology, 43*(1), 238. https://doi.org/10.1037/0012-1649.43.1.238

Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development, 27*(2), 181–194. https://doi.org/10.1016/j.infbeh.2003.09.007

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

Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. USA: Oxford University Press.

J.J. Gibson The senses considered as perceptual systems 1966 Boston: Houghton-Mifflin.

Gliga, T., Bedford, R., Charman, T., Johnson, M. H., Baron-Cohen, S., Bolton, P., … Gammer, I. (2015). Enhanced visual search in infancy predicts emerging autism symptoms. *Current Biology, 25*(13), 1727–1730. https://doi.org/10.1016/j.cub.2015.05.011

Gliga, T., Jones, E. J., Bedford, R., Charman, T., & Johnson, M. H. (2014). From early markers to neuro-developmental mechanisms of autism. *Developmental Review, 34*(3), 189–207. https://doi.org/10.1016/j.dr.2014.05.003

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

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

Habayeb, S., Tsang, T., Saulnier, C., Klaiman, C., Jones, W., Klin, A., & Edwards, L. A. (2021). Visual traces of language acquisition in toddlers with autism spectrum disorder during the second year of life. *Journal of Autism and Developmental Disorders, 51*, 2519–2530. https://doi.org/10.1007/s10803-020-04730-x

Herlihy, L., Knoch, K., Vibert, B., & Fein, D. (2015). Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. *Autism, 19*(1), 20–28. https://doi.org/10.1177/1362361313509731

Hill, E. L., Crane, L., & Bremner, A. J. (2012). Developmental disorders and multisensory perception. In A. Bremner, D. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 273–300). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0012.

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), Article e12381. https://doi.org/10.1111/desc.12381

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development, 76*(3), 598–613. https://doi.org/10.1111/j.1467-8624.2005.00866.x

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and asperger syndrome. *Journal of Autism and Developmental Disorders, 33*(1), 3–13. https://doi.org/10.1023/A:1022270118899

Hudry, K., Chandler, S., Bedford, R., Pasco, G., Gliga, T., Elsabbagh, M., … Charman, T. (2014). Early language profiles in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(1), 154–167. https://doi.org/10.1007/s10803-013-1861-4

Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., & Myowa-Yamakoshi, M. (2019). Audiovisual speech perception and language acquisition in preterm infants: A longitudinal study. *Early Human Development, 128*, 93–100. https://doi.org/10.1016/j.earlhumdev.2018.11.001

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

Iverson, J. M., & Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of Autism and Developmental Disorders, 37*(1), 158–170. https://doi.org/10.1007/s10803-006-0339-z

Jochaut, D., Lehongre, K., Saitovitch, A., Devauchelle, A. D., Olasagasti, I., Chabane, N., … Giraud, A. L. (2015). Atypical coordination of cortical oscillations in response to speech in autism. *Frontiers in Human Neuroscience, 9*, 171. https://doi.org/10.3389/fnhum.2015.00171

Johnson, M. H. (2017). Autism as an adaptive common variant pathway for human brain development. *Developmental Cognitive Neurosciences, 25*, 5–11. https://doi.org/10.1016/j.dcn.2017.02.004

Johnson, M. H., Jones, E. J., & Gliga, T. (2015). Brain adaptation and alternative developmental trajectories. *Developmental Psychopathology, 27*(2), 425–442. https://doi.org/10.1017/S0954579415000073

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

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Science, 2*(10), 389–398. https://doi.org/10.1016/S1364-6613(98)01230-3

Karmiloff-Smith, A. (2015). An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS Neuroscience, 2*(2), 91–104. https://doi.org/10.3934/neuroscience.2015.2.91

Key, A. P., & Stone, W. L. (2012). Processing of novel and familiar faces in infants at average and high risk for autism. *Developmental Cognitive Neuroscience, 2*(2), 244–255. https://doi.org/10.1016/j.dcn.2011.12.003

Kleberg, J. L., Nyström, P., Bölte, S., & Falck-Ytter, T. (2019). Sex differences in social attention in infants at risk for autism. *Journal of Autism and Developmental Disorders, 49*, 1342–1351. https://doi.org/10.1007/s10803-018-3799-z

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

Kushnerenko, E., Tomalski, P., Ballieux, H., Pott on, A., Birtles, D., Frostick, C., & Moore, D. G. (2013). Brain responses and looking behavior during audio-visual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology, 4*(432), 1–10. https://doi.org/10.3389/fpsyg.2013.00432

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1094–1106. https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J. (2010a). Infant perception of audiovisual speech synchrony. *Developmental Psychology, 46*(1), 66. https://doi.org/10.1037/a0015579

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology, 56*(2), 292–315. https://doi.org/10.1002/dev.21197

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

Lewkowicz, D. J., Leo, I., & Simion, F. (2010b). Intersensory perception at birth: newborns match nonhuman primate faces and voices. *Infancy, 15*(1), 46–60. https://doi.org/10.1111/j.1532-7078.2009.00005.x

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule* (Second Edition (ADOS-2). Torrance, CA: Western Psychological Services.

Lozano, I., Belinchón, M., & Campos, R. (2024). Sensitivity to temporal synchrony and selective attention in audiovisual speech in infants at elevated likelihood for autism in the first year: A preliminary longitudinal study. *Infant Behavior and Development, 76*, Article 101973. https://doi.org/10.1016/j.infbeh.2024.101973

Lozano, I., López Pérez, D., Laudańska, Z., Malinowska-Korczak, A., Szmytke, M., Radkowska, A., & Tomalski, P. (2022). Changes in selective attention to articulating mouth across infancy: Sex differences and associations with language outcomes. *Infancy, 27*(6), 1132–1153. https://doi.org/10.1111/infa.12496

Mastergeorge, A. M., Kahathuduwa, C., & Blume, J. (2021). Eye-tracking in infants and young children at risk for autism spectrum disorder: A systematic review of visual stimuli in experimental paradigms. *Journal of Autism and Developmental Disorders, 51*, 2578–2599. https://doi.org/10.1007/s10803-020-04731-w

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology, 56*(2), 154–178. https://doi.org/10.1002/dev.21177

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science, 22*(1), Article e12701. https://doi.org/10.1111/desc.12701

Messinger, D., Young, G. S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., … Hutman, T. (2013). Beyond autism: a baby siblings research consortium study of high-risk children at three years of age. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(3), 300–308. https://doi.org/10.1016/j.jaac.2012.12.011

Miniscalco, C., Fränberg, J., Schachinger-Lorentzon, U., & Gillberg, C. (2012). Meaning what you say? Comprehension and word production skills in young children with autism. *Research in Autism Spectrum Disorders, 6*(1), 204–211. https://doi.org/10.1016/j.rasd.2011.05.001

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

Mullen, E. M. (1995). *Mullen Scales of Early Learning manual*. AGS ed.). American Guidance Service.

Nayar, K., Shic, F., Winston, M., & Losh, M. (2022). A constellation of eye-tracking measures reveals social attention differences in ASD and the broad autism phenotype. *Molecular Autism, 13*(1), 18.

Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., … Steinfeld, M. (2014). The broader autism phenotype in infancy: when does it emerge? *Journal of the American Academy of Child and Adolescent Psychiatry, 53*(4), 398–407. https://doi.org/10.1016/j.jaac.2013.12.020

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., … Hutman, T. (2011). Recurrence risk for autism spectrum disorders: A Baby Siblings Research Consortium study. *Pediatrics, 128*(3), e488–e495. https://doi.org/10.1542/peds.2010-2825

Palomo, R., Ozonoff, S., Young, G., & Belinchón, M. (2022). Social orienting and initiated joint jttention behaviors in 9 to 12 month old children with autism spectrum disorder: A family home movies study. *Autism Research*, 1–11. https://doi.org/10.1002/aur.2695

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry, 52*(5), 588–598. https://doi.org/10.1111/j.1469-7610.2010.02332.x

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

Piccardi, E. S., Begum Ali, J., Jones, E. J., Mason, L., Charman, T., Johnson, M. H., & Gliga, T. (2021). Behavioural and neural markers of tactile sensory processing in infants at elevated likelihood of autism spectrum disorder and/or attention deficit hyperactivity disorder. *Journal of Neurodevelopmental Disorders, 13*(1), 1–18. https://doi.org/10.1186/s11689-020-09334-1

Plumb, A. M., & Wetherby, A. M. (2013). Vocalization development in toddlers with autism spectrum *disorder. Journal of Speech, Language, and Hearing Research, 56*, 721–734. https://doi.org/10.1044/1092-4388(2012/11-0104

Pons, F., Andreu, l, Sanz-Torrent, M., Buil-Legaz, L., & Lewkowicz, D. (2013). Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. *Journal of Child Language, 40*(3), 687–700. https://doi.org/10.1017/S0305000912000189

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science, 26*(4), 490–498. https://doi.org/10.1177/0956797614568320

Pons, F., & Lewkowicz, D. J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica, 149*, 142–147. https://doi.org/10.1016/j.actpsy.2013.12.013

Pons, F., Teixidó, M., Garcia-Morera, J., & Navarra, J. (2012). Short-term experience increases infants' sensitivity to audiovisual asynchrony. *Infant Behavior and Development, 35*(4), 815–818. https://doi.org/10.1016/j.infbeh.2012.06.006

Poole, D., Gowen, E., Poliakoff, E., Lambrechts, A., & Jones, L. A. (2023). When 2 become 1: Autistic simultaneity judgements about asynchronous audiovisual speech. *Quarterly Journal of Experimental Psychology, 77*(9), 1865–1882. https://doi.org/10.1177/17470218231197518

Quigley, J., McNally, S., & Lawson, S. (2016). Prosodic patterns in interaction of low-risk and at-risk-of-autism spectrum disorders infants and their mothers at 12 and 18 months. *Language Learning and Development, 12*(3), 295–310. https://doi.org/10.1080/15475441.2015.1075405

Riboldi, E. M., Capelli, E., Cantiani, C., Beretta, C., Molteni, M., & Riva, V. (2024). Differentiating early sensory profiles in toddlers at elevated likelihood of autism and association with later clinical outcome and diagnosis. *Autism, 28*(7), 1654–1666. https://doi.org/10.1177/13623613231200081

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

Riva, V., Riboldi, E. M., Dondena, C., Piazza, C., Molteni, M., & Cantiani, C. (2022). Atypical ERP responses to audiovisual speech integration and sensory responsiveness in infants at risk for autism spectrum disorder. *Infancy, 27*(2), 369–388. https://doi.org/10.1111/infa.12456

Santapuram, P., Feldman, J. I., Bowman, S. M., Raj, S., Suzman, E., Crowley, S., … Woynaroski, T. G. (2022). Mechanisms by which early eye gaze to the mouth during multisensory speech influences expressive communication development in infant siblings of children with and without autism. *Mind, Brain, and Education, 16*(1), 62–74. https://doi.org/10.1111/mbe.12310

Schaeffer, J., Abd El-Raziq, M., Castroviejo, E., Durrleman, S., Ferré, S., Grama, I., … Tuller, L. (2023). Language in autism: Domains, profiles and co-occurring conditions. *Journal of Neural Transmission, 130*, 433–457. https://doi.org/10.1007/s00702-023-02592-y

Schormans, A. L., & Allman, B. L. (2023). An imbalance of excitation and inhibition in the multisensory cortex impairs the temporal acuity of audiovisual processing and perception. *Cerebral Cortex, 33*(18), 9937–9953. https://doi.org/10.1093/cercor/bhad256
Schultz, R. T. (2005). Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience, 23*(2-3), 125–141. https://doi.org/10.1016/j.ijdevneu.2004.12.012
Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biological Psychiatry, 75*(3), 231–237. https://doi.org/10.1016/j.biopsych.2013.07.009
Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J., & Lewkowicz, D. J. (2012). The development of audiovisual speech perception. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 207–228). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0009.
Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*(6), 626. https://doi.org/10.1037/0012-1649.15.6.626
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., 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
Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience, 34*(3), 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014
Suri, K. N., Whedon, M., & Lewis, M. (2023). Perception of audio-visual synchrony in infants at elevated likelihood of developing autism spectrum disorder. *European Journal of Pediatrics*, 1–13. https://doi.org/10.1007/s00431-023-04871-y
Szatmari, P., Bryson, S. E., Boyle, M. H., Streiner, D. L., & Duku, E. (2003). Predictors of outcome among high functioning children with autism and Asperger syndrome. *Journal of Child Psychology and Psychiatry, 44*(4), 520–528. https://doi.org/10.1111/1469-7610.00141
Tager-Flusberg, H. (2010). The origins of social impairments in autism spectrum disorder: Studies of infants at risk. *Neural Networks, 23*(8-9), 1072–1076. https://doi.org/10.1016/j.neunet.2010.07.008
Tager-Flusberg, H., Paul, R., & Lord, C. E. (2005). Language and communication in autism. In F. Volkmar, R. Paul, A. Klin, & D. J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorder* (3rd ed., pp. 335–364). New York: Wiley.
Taylor, N., Isaac, C., & Milne, E. (2010). A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *Journal of Autism and Developmental Disorders, 40*(11), 1403–1411. https://doi.org/10.1007/s10803-010-1000-4
Thye, M. D., Bednarz, H. M., Herringshaw, A. J., Sartin, E. B., & Kana, R. K. (2018). The impact of atypical sensory processing on social impairments in autism spectrum disorder. *Developmental Cognitive Neuroscience, 29*, 151–167. https://doi.org/10.1016/j.dcn.2017.04.010
Todd, J. T., & Bahrick, L. E. (2022). Individual differences in multisensory attention skills in children with autism spectrum disorder predict language and symptom severity: Evidence from the multisensory attention assessment protocol (MAAP). *Journal of Autism and Developmental Disorders, 53*(12), 4685–4710. https://doi.org/10.1007/s10803-022-05752-3
Treille, A., Cordeboeuf, C., Vilain, C., & Sato, M. (2014). Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions. *Neuropsychologia, 57*, 71–77. https://doi.org/10.1016/j.neuropsychologia.2014.02.004
Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology, 169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002
de Villiers Rader, N., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: the importance of dynamic synchrony. *Language Sciences, 34*(5), 559–568. https://doi.org/10.1016/j.langsci.2012.03.011
Viswanathan, N. K., de Klerk, C. C., Wass, S. V., & Goupil, L. (2024). Learning to imitate facial expressions through sound. *Developmental Review, 73*, Article 101137. https://doi.org/10.1016/j.dr.2024.101137
Walker-Andrews, A. S. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz, & R. Lickliter (Eds.), *The development of intersensory perception* (pp. 39–56). Psychology Press.
Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64*, 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005
Wallace, M. T., Woynaroski, T. G., & Stevenson, R. A. (2019). Multisensory integration as a window into orderly and disrupted cognition and communication. *Annual Review of Psychology, 71*, 193–219. https://doi.org/10.1146/annurev-psych-010419-051112
Weiland, R. F., Polderman, T. J., Smit, D. J., Begeer, S., & Van der Burg, E. (2023). No differences between adults with and without autism in audiovisual synchrony perception. *Autism, 27*(4), 927–937. https://doi.org/10.1177/13623613221121414
Whitehouse, A. J., Barry, J. G., & Bishop, D. V. (2007). The broader language phenotype of autism: A comparison with specific language impairment. *Journal of Child Psychology and Psychiatry, 48*(8), 822–830. https://doi.org/10.1111/j.1469-7610.2007.01765.x
Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(12), 2891–2902. https://doi.org/10.1007/s10803-013-1836-5
Yirmiya, N., Gamliel, I., Pilowsky, T., Feldman, R., Baron-Cohen, S., & Sigman, M. (2006). The development of siblings of children with autism at 4 and 14 months: Social engagement, communication, and cognition. *Journal of Child Psychology and Psychiatry, 47*(5), 511–523. https://doi.org/10.1007/s10803-006-0329-1
Yirmiya, N., & Ozonoff, S. (2007). The very early autism phenotype. *Journal of Autism and Developmental Disorders, 37*(1), 1–11. https://doi.org/10.1007/s10803-006-0329-1
Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science, 12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x
Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., … Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research, 15*, 778–790. https://doi.org/10.1002/aur.2696
Zerr, M., Freihorst, C., Schütz, H., Sinke, C., Müller, A., Bleich, S., Münte, T. F., & Szycik, G. R. (2019). Brief sensory training narrows the temporal binding window and enhances long-term multimodal speech perception. *Frontiers in Psychology, 10*, 2489. https://doi.org/10.3389/fpsyg.2019.02489
Zhao, Z., Tang, H., Zhang, X., Zhu, Z., Xing, J., Li, W., … Lu, J. (2023). Characteristics of visual fixation in Chinese children with autism during face-to-face conversations. *Journal of Autism and Developmental Disorders, 53*(2), 746–758. https://doi.org/10.1007/s10803-021-04985-y
Zhou, H. Y., Cui, X. L., Yang, B. R., Shi, L. J., Luo, X. R., Cheung, E. F., … Chan, R. C. (2022). Audiovisual temporal processing in children and adolescents with schizophrenia and children and adolescents with autism: evidence from simultaneity-judgment tasks and eye-tracking data. *Clinical Psychological Science, 10*(3), 482–498. https://doi.org/10.1177/21677026211031543