brain sciences

MDPI

Review

The Role of Talking Faces in Infant Language Learning: Mind the Gap between Screen-Based Settings and Real-Life Communicative Interactions

Joan Birulés [1,*], Louise Goupil [1], Jérémie Josse [1] and Mathilde Fort [1,2]

[1] Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105, Université Grenoble Alpes, 38058 Grenoble, France; louise.goupil@univ-grenoble-alpes.fr (L.G.); jeremie.josse@univ-grenoble-alpes.fr (J.J.); mathilde.fort@univ-grenoble-alpes.fr (M.F.)

[2] Centre de Recherche en Neurosciences de Lyon, INSERM U1028-CNRS UMR 5292, Université Lyon 1, 69500 Bron, France

* Correspondence: joan.birules@univ-grenoble-alpes.fr

Abstract: Over the last few decades, developmental (psycho) linguists have demonstrated that perceiving talking faces audio-visually is important for early language acquisition. Using mostly well-controlled and screen-based laboratory approaches, this line of research has shown that paying attention to talking faces is likely to be one of the powerful strategies infants use to learn their native(s) language(s). In this review, we combine evidence from these screen-based studies with another line of research that has studied how infants learn novel words and deploy their visual attention during naturalistic play. In our view, this is an important step toward developing an integrated account of how infants effectively extract audiovisual information from talkers' faces during early language learning. We identify three factors that have been understudied so far, despite the fact that they are likely to have an important impact on how infants deploy their attention (or not) toward talking faces during social interactions: social contingency, speaker characteristics, and task- dependencies. Last, we propose ideas to address these issues in future research, with the aim of reducing the existing knowledge gap between current experimental studies and the many ways infants can and do effectively rely upon the audiovisual information extracted from talking faces in their real-life language environment.

Keywords: talking faces; audiovisual speech perception; infancy; language acquisition; naturalistic interactions; audiovisual speech cues

check for updates

Citation: Birulés, J.; Goupil, L.; Josse, J.; Fort, M. The Role of Talking Faces in Infant Language Learning: Mind the Gap between Screen-Based Settings and Real-Life Communicative Interactions. *Brain Sci.* **2023**, *13*, 1167. https://doi.org/10.3390/brainsci13081167

Academic Editor: Kaisa Tiippana

Received: 10 July 2023
Revised: 28 July 2023
Accepted: 1 August 2023
Published: 5 August 2023

1. Introduction

Speech is typically perceived multimodally: most social interactions occur face to face, so we not only hear our interlocutor but also see the information inherent to any facial articulatory movement and inherent to linguistic communication more specifically. Primarily, through the eye's region, faces convey a good amount of information about a person's state of mind, attitude, and potential intentions, as well as referential information and speech rhythms (see [1] for a review). Additionally, through the talker's mouth region, we gain access to spatiotemporal and acoustic congruent auditory and visual speech cues [2,3]. Given that auditory and visual cues provide overlapping information about the same speech event, they have often been named "redundant audiovisual (AV) speech cues" (e.g., [4,5]). Prior studies have shown that access to such AV redundant speech cues, compared to auditory-only situations, can facilitate lexical access and, more largely, speech comprehension, most notably when the acoustic signal becomes difficult to understand due to noise [6–11] or to an unfamiliar accent or language (e.g., [12–14]). In such occasions, adult listeners have been shown to increase their visual attention (hereafter attention) to the talker's mouth in order to maximize the processing of AV speech cues and enhance their processing of speech; for instance, when background acoustic noise increases [15,16], when volume is low [17], when their language proficiency is low [18–20], or when they are performing particularly challenging speech-processing tasks (e.g., speech segmentation [21] or sentences comparison [18]). On the other hand, when speech-processing demands are reduced, adults modulate their attention and focus more on the eyes of the talker [21–23], which can also support language understanding by constraining interpretations (e.g., of the speaker's current emotion or current focus of attention).

Overall, these studies suggest that adults flexibly modulate their selective attention to subparts of a talking face as a function of factors such as the task at hand, the amount of noise in the sensory environment, and language experience. If adults do this to maximize the processing of redundant AV speech cues inherent in talkers' faces and help them process speech, we might hypothesize that infants could also take advantage of such cues when learning their first language/s.

In the present article, we review existing evidence on whether and how the infant's visual perception of dynamic talking faces is linked to language learning. First, we do so by describing findings from prior screen-based studies on infant face perception, including selective attention to talking faces and its relation to language learning (Section 1). Second, we present evidence from naturalistic studies on infants' perception and attention to their caregiver's faces during play and discuss the apparent differences in looking patterns towards talking faces that have been observed between screen-based and real-life studies. Third, we describe three variables (i.e., social contingency, speaker-dependency, and task-dependency) that we believe are currently understudied and should be further explored to better understand how infants dynamically and flexibly exploit talking faces to learn their native language/s (Section 3). Finally, we present our conclusions and directions for future studies (Section 4) and propose several ideas to reach a more complete knowledge of the role that speakers' talking faces play in infants' language acquisition.

2. The Role of Faces in Infants' Attention and Language Acquisition: Evidence from Screen-Based Settings

2.1. Infants' Preference for Faces

During their first year of life, most infants' direct communicative interactions occur face to face with their caregivers (for an estimation in ecological situations, see [24,25]), which allows them to gain access to rich AV cues that dynamic talking faces provide. Crucially, on top of the fact that faces are readily accessible in infants' typical visual field, screen-based studies attempting to mimic these face-to-face situations in the lab have revealed that infants also show a preference for face-like stimuli from very early on: they tend to orient toward faces preferably when in competition with other stimuli (e.g., [26]). It has been suggested that this early "face bias" is already present in newborns, who look longer on face-like patterns than inverted face-like patterns or different geometric patterns [27–29], although it is unclear whether this reflects a genuine preference for "faces" per se rather than a preference for some lower-level visual features of faces. Later, once infants' visual system begins to be fully functional around 3 to 4 postnatal months [30], and they have gained some experience with their environment, infants' preference for faces continues to increase, particularly when presented amongst multiple competing objects (4- to 12-month-old infants [31,32]). These studies suggest that infants routinely have a direct visual access to talking faces, and that they show a very early interest and increased attention to such information. Other studies suggest that infants' attentional preference for faces after the age of 3–4 months cannot entirely be explained by low-level visual features, reflecting the onset of an active search for relevant information in faces [31]. In line with this hypothesis, studies have also shown that infants do not attend to speakers' faces as much when they do not engage in communicatively relevant interactions, such as a mutual gaze [33,34], which enhances their cortical processing of faces and face recognition skills [35,36].

In summary, it has been discussed that learning to process interlocutors' faces is likely to be a crucial aspect of infants' social [37,38] and language learning (for recent reviews, see: [39–41]). However, which specific facial cues (i.e., eyes and eye-gaze, AV speech cues from the talker's mouth) infants attend to and exploit at each moment in development and for which type of learning they might do so remains to be understood.

2.2. Infant' Attention to the Eyes and Mouth of a Talking Face, and Its Relation to Language Acquisition

Prior studies have approached this question by exploring infants' selective attention to videos of talking faces, analyzing their looking time to the internal features of a talker's face (i.e., eyes and mouth) across development. Furthermore, some have examined how such measures of attention relate to infants' concurrent and later language outcomes, providing correlational evidence that attending to the eyes or the mouth region of a talking face at different stages of development can support language learning (e.g., [42–48]).

On the one hand, studies suggest that paying attention to the eye region of a face can boost infant language learning: a direct gaze—rather than an indirect gaze—establishes the intent to communicate, and young infants are sensitive to these communicative signals [33,34,49,50]. Later, infants develop gaze- and head-following skills, which allow them to orient their attention at the same location as their social partners, creating joint attention moments between the infant and the caregiver, which can support word learning (see [41,51] for recent overviews). Of particular relevance for lexical acquisition, following the gaze/head of a speaker in the direction of a named object can allow listeners to disambiguate which object is associated with the novel spoken label [52], thus solving the referential uncertainty problem (i.e., the uncertainty inherent in mapping a heard name to its intended referent, in a complex and variable environment [53]). In line with this idea, several studies have shown that gaze-following behavior in the first year is positively correlated with vocabulary development in the second year [43,54–56]. Additional experimental evidence suggests that infants' encoding of the visual property of a novel object can be increased when they look toward it by following another person's gaze [35]. Moreover, gaze following also supports infants' mapping of novel words to the object that the speaker is looking at [41,57].

On the other hand, paying attention to the mouth of a talking face can also be useful for enhancing speech processing and infant language acquisition (e.g., [58]). As reviewed above, adults rely on the AV speech cues of a talker's mouth to enhance speech perception, especially when speech processing becomes challenging (e.g., [15]). Prior evidence shows that infants are sensitive to AV correspondences from 2 to 3 months of age (e.g., [59,60]). Their auditory perception of speech is influenced by its concurrent visual presentation from 5 months of age (i.e., McGurk effect; [61]), and they can use AV speech information to discriminate and learn speech sounds [62,63], segment speech [64] and increase word processing and retention [65,66]; see for reviews [39,41]. It is then reasonable to hypothesize that orienting and sustaining one's attention on a talker's mouth could be a good strategy for enhancing language acquisition at various levels during infancy.

Lewkowicz and Hansen-Tift [42] performed the first study to explore the relationship between infants' selective attention to a video of a talker's face (i.e., to the eyes and mouth) across their first year of life and its relation to language processing [42]. To do so, the authors recorded the gaze of 4- to 12-month-old monolingual English-learning infants whilst watching a speaker talk in their native language. The results first showed that similar to newborns and younger infants [26,67], 4-month-old infants attended more to the talker's eyes. Crucially, however, infants then started shifting their attention toward the mouth of the talking face, showing equal attention to the eyes and mouth at 6 months and more attention to the mouth from 8 months of age [42].

This attentional shift from the eyes to the mouth of the talker was interpreted by the authors as evidence of infants' emerging interest in processing AV speech. Given that 6 months of age is also the stage where infants' endogenous control starts to emerge [68,69], this would allow them to exert top-down control to specific subparts of a speaker's face and voluntarily deploy more of their attention into the speaker's mouth. Indeed, the onset of this mouth preference has been confirmed to emerge at around 6–8 months of age by more recent studies [44–47,70–73] and seems to present its peak at around 18 months of age [46,74]. This disposition to look toward speakers' mouths remains present in the second year [75,76] and slowly diminishes during later childhood, with 5-year-old children typically showing balanced attention between the talker's eyes and mouth when perceiving native speech [46,77,78].

Overall, this suggests that infants and children deploy an important amount of their selective attention to the mouth of talking faces during their first years of life and that such preference for the mouth diminishes with increasing age and language proficiency. However, to what extent are infants' attentional patterns to the subparts of talking faces reflective of their interest in the processing of AV speech cues (rather than being determined exogenously as a function of relative saliency)? Additionally, is this attentional strategy efficient in helping them learn their native language/s?

Recent studies have provided empirical evidence supporting the link between attention to the mouth and different aspects of language processing. For instance, studies have shown that infant-directed speech (IDS)—which is known to enhance infant attention to speech, but also language processing and learning [79]—elicits greater attention to the mouth during the first and second year of life [75,80,81]. The fact that IDS induces infants to attend more to the mouth of talkers suggests that this may be an efficient strategy of IDS to increase the processing of AV speech cues and, in turn, boost infants' processing of speech. In a similar vein, a recent study has shown that 15-month-old infants also increased their attention to a talker's mouth as they learned new non-adjacent rules from unfamiliar speech [82]. This has been interpreted as reflecting infants' additional reliance on AV speech cues to help process and acquire the syntactic rules of their native language.

Additional evidence comes from studies showing that similar to adults, infants also increase their attention to the mouth of a talker when speech processing becomes more challenging, as is the case of perceiving non-native speech [42,47,76,83] (although not in [46]) or growing up in bilingual environments and learning two rhythmically and phonologically close languages, such as Spanish and Catalan. In the latter case, close-language bilingual infants deploy more of their attention to a talker's mouth than monolingual infants or bilingual infants learning phonologically distant languages [47,76,84,85], which continues into later childhood [76,86] and is likely to enhance language differentiation and processing under a linguistically more challenging environment. Taken together, these findings suggest that, like children and adults, infants already adapt their looking patterns to subparts of a talking face flexibly and deploy more or less attention to the AV speech cues of a talker's mouth as a function of local demands concerning speech processing and language learning.

Crucially, some studies have also provided direct correlational evidence showing an association between infants' increased attention to the mouth and the different aspects of language learning. For instance, Young and colleagues [73] recorded 6-month-olds' eye gaze in a video of a live mother–infant interaction and revealed that greater mouth-looking predicted higher expressive language at 18 months. In the same vein, Tenenbaum and colleagues [43] showed that looking time at a talker's mouth as well as gaze-following behavior at 12 months of age predicted infants' vocabulary size at 18 and 24 months of age. Using a different paradigm, Imafuku and colleagues [48] presented 6-month-old infants with videos of a speaker producing vowel sounds, while recording their eye gaze and vocal productions. The results showed that infants vocalized more when their faces were upright, made eye contact, and when infants looked more at the speakers' mouths [48]. The correlation between mouth-looking and language skills has also been shown by other studies, albeit with differences in age and the specific aspects of lexical development (see [44,45]).

To summarize, it emerges from these findings that, over time, infants increasingly deploy their attention toward the AV speech cues of a talker's eyes and mouth regions to help them process and learn their native language/s. However, it is worth noting that these conclusions originate from screen-based studies. Such experimental situations generally differ from real-world interactions (for a recent overview, see: [87]) and, therefore, we must also explore whether and how these mechanisms are effectively used and relied upon by infants in their complex and culturally situated language environment, and during free-flowing, dynamic social interactions like proto-conversations, play, etc. In the following section, we thus explore infants' selective attention to caregivers' faces in real-life language learning situations.

3. The Role of Faces in Infants' Attention and Language Acquisition: Evidence from Real-Life Communicative Interactions

During real-life interactions, talking faces appear within the realm of complex, multimodal, and only partially predictable socio-communicative interactions. This is a different situation compared to what has happened in most studies examining the role of talking faces for infant language learning using screen-based tasks. Indeed, in virtually all these studies (including some work from the authors), non-contingent face stimuli have often appeared exaggerated in size and presented against a rather neutral background with no or very few distractors on a computer screen (e.g., [42,44,46,70,72,74,76,84]). These presentation characteristics could lead to a distortion in the perceptual saliency of talking faces and their role in guiding infants' attention and language processing and learning.

Bahrick and colleagues posited that redundant and temporally synchronized AV cues, such as the ones provided by the talking mouths of speakers, "pop out" from the background and other less salient distractors, directing attentional selectivity (i.e., the Intersensory Redundancy Hypothesis, IRH, [88,89]). In other words, they suggest that talking faces offer a form of pre-attentive bottom-up saliency that would automatically attract the perceiver's visual attention when surrounded by other less salient distractors. This idea has been previously challenged, however, by studies in adults showing that the detection of temporally synchronized and coherent AV cues is not a pre-attentive process given that it is sensitive to attentional demands [90–92] and that it requires a top-down serial time search, both in adults [93] and in children [94].

Another prediction that derives from IRH is that talking faces should be highly and frequently looked at when present in an infant's environment. This preference for talking faces should be observed from around 2 to 3 months of age once infants become sensitive to AV correspondences [59,60]. At this developmental stage, AV correspondences should make faces more salient and catch infants' attention in a bottom-up fashion. As cognitive control matures with age, the preference for talking faces should increase, caused by both the bottom-up system, responding to the highly salient features of AV speech, and by gradually developing top-down forms of endogenous attention [68,95], assuming that infants endogenously seek the social and linguistic information that talking faces provide.

Screen-based studies support this hypothesis by showing that infants prefer to look at silent static or talking faces rather than objects in their first months of life [28,96,97] and that crucially, this preference increases with age in typically developing infants [31,32,98,99] (although see [97] for an age decrease) independent of the low-level saliency features of faces [31,98]. In such screen-based studies, when exploring the role of talking faces in early language acquisition, the estimated percentage of looking time to faces has been reported between 30 and 90%, depending on age and stimulus features [31,32,98–100].

Differently, however, if we extract descriptive statistics from real-life studies that examine how infants visually explore their environment in naturalistic contexts, which are thought to afford language learning (e.g., joint play), data suggest that faces comprise a smaller percentage of infants' visual input (i.e., between 10 and 30%), and that such a percentage seems to decrease with age [24,25]. For instance, in free toy play situations, infants from 11 to 12 months of age achieve joint attention moments with their caregivers toward objects that are close to them (e.g., [101]) with little attentional guidance from their caregiver's faces and head direction [102–105], and a seemingly greater impact of manual gestures [106]. These data could suggest that faces play a minor role in children's social and language learning, at least after 11–12 months of life. It is expected that in free-flowing interactions, with higher stimulation and competition, infants should show lower face-looking times than in screen-based studies. Yet, this should remain constant over time; therefore, the fact that this preference decreases with age remains to be explained, as it contradicts the increase in age shown in screen-based studies. In other words, collectively, these findings suggest that the perceptual saliency of faces in naturalistic settings might be lower than what we would predict based on screen-based studies and the IRH [89].

Relatedly, infants' free vs. restrained movement reasonably have a strong impact on infant attention. As infants grow up, motor and postural developments progressively transform their visual input, increasing its complexity and diversity [107,108] and affecting the relative saliency of talking faces. Infants shift from laying on their back and mainly looking up—towards the ceiling and the face(s) of their caregiver(s)—to sitting, then standing up and walking. Such novel postural and motor control affords infants a novel perspective of their environment by looking more frontally, reaching and grasping objects that surround them. Under this rationale, directly looking toward adults' faces becomes less frequent, and infants engage more with specific subsets of their visual environment (e.g., hands, objects, food, etc., see [108]).

Overall, these studies exploring naturalistic interactions where infants move freely suggest that faces may be perceptually less salient and less frequently attended to than previously assumed based on screen-based studies and that, with increasing age, infants' attention to faces diminishes. This suggests that infants' need to deploy their attention toward the AV speech cues of speakers' faces may also gradually decrease with increasing age and language proficiency. A potential explanation is that, indeed, infants' information-seeking behavior becomes more and more optimal as they grow up and that, therefore, given that their cognitive and linguistic abilities improve, their selective attention to faces becomes less frequent but more targeted to certain specific situations when AV speech cues become relevant. These may involve situations where communicative information becomes unclear (due to referential uncertainty, speech ambiguity, noise, L2 speech, etc.) or when other social and emotional factors come into play (face identification, emotional reassurance, etc.). Therefore, infants may prioritize other stimuli while communication is clear and fluent and only look toward speakers' faces in ambiguous situations.

In our view, these discrepancies might also arise from the fact that three factors have been insufficiently considered in past studies, even though they systematically vary by context and might deeply impact the results that are observed in any given context. In the following section, we provide the reader with an overview of these three factors that we believe may influence infants' attention and the use of talking faces in naturalistic settings, which are currently understudied. In turn, we also discuss how these factors may also help explain the differences in attention to faces between the screen-based and real-life settings above described.

4. Moving Closer to Real-Life Language Learning Interactions: Three Factors That Deserve More Attention

4.1. The Influence of Dyadic Bidirectional Contingency

The first factor that we consider to be understudied in screen-based studies exploring the role of talking faces in language learning—that has also not been systematically explored or manipulated in naturalistic studies—concerns the bidirectional contingency between the infant and the talking face the infant is looking at. Typically, screen-based approaches provide very little temporal contingency between the shown stimuli and the infant's behavior and mainly focus on the unidirectional adaptation of the infant to talking faces rather than considering this relation in a bidirectional fashion (e.g., [42,46,48,76]). In other words, in most screen-based paradigms, the behavior of the talking face is determined by the experimenter (e.g., pre-recorded) rather than by the participant. Talking faces almost never adapt to infants' attentional variations across time (e.g., gaze direction), including emotional or sociolinguistic responses (e.g., smile, pointing, presence of babbling, or imitation of the speaker). Overall, this lack of bidirectional contingency could lead to an underestimation of the dynamic components of dyadic interactions in real-life situations, which are known to modulate infants' attention and language learning, e.g., [101,109–111]. More specifically, it could lead to an overestimation of the infants' adaptation to the caregiver's face behavior and/or underestimating the adaptation of the caregiver's face as a response to the infant's behavior.

Studies examining real-life interactions between infants and their parents (e.g., during joint play) have shown large inter-individual variability in the extent to which caregivers tend to contingently respond to their infants' behaviors (a measure referred to as contingency, sensitivity, or responsiveness), and this seems to be a unique and important predictor for infant word learning [112–114]. For instance, the more caregivers show temporally contingent responsivity to their infant's behavior, the more their infant is attentive [115–117], learns novel rules [118], novel words [112–114,119], remains sensitive to non-native phonetic contrasts [120–122], and produces more mature speech-like vocalizations [111]. The importance of social contingency for learning has been suggested to stem from the fact that it allows infants to better predict the consequences of their actions and, as such, to better connect causes and consequences during social interactions. This is thought to be key for them to progressively understand that specific behaviors (words, gestures, etc.) are referential (i.e., they are caused by specific mental representations possessed by senders), which can then support a progressive internalization of their meanings [113,114,123].

In summary, the growing literature suggests that social contingency is a key factor for infant learning, including language acquisition, but the extent to which it also shapes the way infants attend to and process a speaker's face remains unclear. A necessary next step for research on the use of AV speech cues for infants' language learning will, thus, be, in our opinion, to examine how social contingency and other factors related to the dialogic nature of early caregiver–infant communication (e.g., turn-taking rates and structure, the predictability of social interactions, etc. [124,125]), shape infants' disposition to selectively attend to and process the speakers' faces, eyes and mouths.

4.2. The Influence of Speakers' Characteristics

Another variable that typically differs from screen-based to real-life studies, and which has also been largely understudied, is the type of speakers that infants are presented with or interact with. Screen-based settings that measure infant attention to talking faces systematically use unrelated strangers that often talk in IDS. By contrast, in real-life settings, infants are typically interacting with their caregivers, who sometimes show less consistent uses of IDS, for instance, when infants are not responsive to their solicitations (e.g., [126]). It is, therefore, possible that a speaker's familiarity and differential use of IDS could account for the differences in attention to faces observed between screen-based and real-life setting studies (see Section 3).

First, low familiarity with an unknown speaker could enhance infants' attention to the speaker's face and talking mouth: such a strategy could help infants build speech representations across speakers (i.e., speaker normalization): a necessary and non-trivial step in early language acquisition [127,128]. Paying more attention to the caregiver's face could also help infants evaluate the expertise of the unfamiliar speaker, estimating, for instance, their social group. Indeed, categorizing the speaker as either belonging or not to the same social group as the infant (in or out group) can change the way they expect and process speech information from this person (see [129,130]).

Second, the more systematic use of IDS in screen-based than in real-life settings could also influence the amount of time infants spend looking toward the face and, more specifically, the talking mouth of their caregivers. Indeed, screen-based settings that experimentally manipulate this factor have shown that IDS is more often associated with faces [131,132] and elicits more mouth looks than adult-directed speech [75,80,81]. In other words, the salient acoustic features of IDS could help infants orient and sustain not only their auditory attention to the speech signal (see [133] for a compatible hypothesis) but also their visual attention to the face, and especially to the talking mouth of their caregivers. In line with this idea, caregivers have been shown to flexibly modify the pitch of their IDS as a function of their infant's responsiveness/feedback [126]. Further research should study more precisely the acoustic features that might attract infants' visual attention toward the face or mouth of speakers and whether this behavior is also observed in more complex and usually noisier real-life situations.

4.3. The Influence of Dyadic Bidirectional Contingency

The last variable that we believe deserves attention, and that also differs from screen-based to real-life studies, is the type of interaction or situation that infants are presented with during an experiment. Most of the evidence gathered from screen-based studies originates from one type of situation: a video showing a close-up of a speaker's face, talking to the infant, against a neutral background, with no or a very small number of objects or other types of distractors. This situation is supposed to mimic everyday face-to-face interactions where parents talk to their infant while sitting on a chair (e.g., during eating, spoon-feeding) or laying down on his/her back (e.g., during a diaper change or before sleeping). During the first months of their life, while mobility is reduced, real dyadic interactions can resemble these—where faces are at a short distance—and observations of caregivers and their infants younger than 9–12 months reflect this [24,25]. However, once infants are older and start to determine their visual inputs mostly by themselves, their experiences become more complex (see Section 2.1), and, therefore, screen-based studies fail to capture infants' most usual types of interaction. On the other hand, naturalistic studies have typically explored how infants deploy their attention in situations of free play with various toys, often in the context of a situation where caregivers and their infants sit opposite to each other around a table, hence diminishing the relative size and saliency of faces in comparison to most screen-based settings [24,25].

These situational differences are relevant and should be considered when drawing conclusions about how infants deploy their attention, as infants' attentional strategies are known to be task-dependent [100,134,135], and, therefore, infants adapt and change their behavior according to the typology of these situations. Prior studies using head-mounted eye trackers have shown that, indeed, infants mostly modulate their attention to the body and face of their mother as a function of task/situation (e.g., reaching, removing obstacles, crawling) and the mother's position and actions [136].

Further research should thus explore the dynamics of infants' selective attention across various situations (e.g., face-to-face interaction during feeding or table-top play sessions vs. side-by-side play interactions on a mat, etc.) in order to understand the way in which these specific settings (i.e., situational or task demands) modulate infants' reliance on the AV speech cues of talking faces.

5. Conclusions and Directions for Future Research

In the present paper, we discuss the role of infants' attention to talking faces as a mechanism to enhance speech processing and language learning (see also [39,40] for screen-based studies reviews). The originality of our approach is that it puts into perspective findings from both screen-based and real-life settings and then focuses on the specific factors that might be most influential yet are currently understudied in both methodologies that explore AV speech perception for early language acquisition.

While it appears clear that attending and processing AV speech cues from talkers' faces is an important mechanism that supports infants' language learning, we have seen that studies to date remain quite specific to their context and methodology (e.g., screen-based, non-contingent paradigms, or real-life, observational-only paradigms, all providing mostly correlational evidence) and that there are several important factors that need to be better explored in order to reduce the gap between our experimental studies and infants' real-life learning mechanisms.

Crucially, given that infants' attention to faces in real-life interactions is less frequent than previously thought or observed by screen-based studies and that it diminishes with increasing age, it becomes highly relevant for future studies to identify in which specific situations of their daily life, such as from the perspective of language learning, infants and children seek AV speech information in their caretaker's faces. To approach this question, we might have to explore factors pertaining to the specific task or experimental situation, resembling real-life routines (e.g., play, book sharing, eating) that are known to foster language acquisition, but also factors relative to the person addressing the infant (e.g., caregiver, unknown experimenter, foreign language speaker, etc.). Finally, we believe that the dynamic aspects of communication, such as social contingency (i.e., the adaptation of infants' behavior to the caregiver's face and vice versa), should also be explored in future studies, aiming to move away from "passive-learning" paradigms and introduce contingency and infants' intrinsic motivation to move closer to real-life learning situations.

Given that infants allocate their attention as a function of their current informational needs and goals [137–139], it is likely that the above-mentioned factors (i.e., speakers' familiarity, task dependencies, and social contingency) change infants' motivation to learn from and about speakers [140], thus modulating their attention toward talking faces, and, in turn, learning. Future studies that systematically vary the relevance of speakers, faces, and contrasting stimuli using different methodologies could help to better understand the extent to which infants' early preference for faces is perceptual and category-specific rather than attention-based and dependent on self-relevance and motivational factors.

Finally, this perspective also suggests that inter-individual differences in infants' typical home environment and caregiving practices could greatly modulate infants' attention and reliance on talking faces for learning their native language, for instance, as a function of the relative frequency of face-to-face vs. side-to-side interactions in their homes. Most research in this area has focused on Westerners, so it remains plausible that exploiting visual information from talking faces is only one of many strategies that infants could use to support language acquisition. For instance, face-to-face interactions involving a mutual gaze are thought to be more common in Western cultures compared to side-to-side interactions involving mutual touch in other cultures [141]. It is, therefore, possible that other strategies and modalities (e.g., touch, gestures) are more privileged than the visual modality outside of Western cultural settings. Studying infants' attention to talking faces and bodies in culturally more varied samples and real-life interactions is, therefore, an important venue for future research and is important to better document diversities and universalities in the ways infants learn their native language/s.

Author Contributions: Conceptualization, J.B., L.G., J.J. and M.F.; writing—original draft preparation, J.B., L.G. and M.F.; writing—review and editing, J.B., L.G., J.J. and M.F.; supervision, M.F.; project administration, M.F.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NeuroCog, the Pole Grenoble Cognition, the SFR Santé Société and the Fondation de France (Grant number: 00112028).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Birmingham, E.; Kingstone, A. Human Social Attention. *Ann. N. Y. Acad. Sci.* **2009**, *1156*, 118–140. [CrossRef] [PubMed]
2. Chandrasekaran, C.; Trubanova, A.; Stillittano, S.; Caplier, A.; Ghazanfar, A.A. The Natural Statistics of Audiovisual Speech. *PLoS Comput. Biol.* **2009**, *5*, e1000436. [CrossRef]

3. Yehia, H.; Rubin, P.; Vatikiotis-Bateson, E. Quantitative association of vocal-tract and facial behavior. *Speech Commun.* **1998**, *26*, 23–43. [CrossRef]
4. Bahrick, L.E.; Lickliter, R. Intersensory redundancy guides early perceptual and cognitive development. In *Advances in Child Development and Behavior*; Elsevier Masson SAS: Amsterdam, The Netherlands, 2003; pp. 153–187. [CrossRef]
5. Lewkowicz, D.J. Infant perception of audio-visual speech synchrony. *Dev. Psychol.* **2010**, *46*, 66–77. [CrossRef]
6. Cotton, J.C. Normal 'Visual Hearing. *Science* **1935**, *82*, 592–593. [CrossRef]
7. Risberg, A.; Lubker, J. Prosody and speechreading. *Q. Prog. Status Rep.* **1978**, *4*, 1–16.
8. Sumby, W.H.; Pollack, I. Visual Contribution to Speech Intelligibility in Noise. *J. Acoust. Soc. Am.* **1954**, *26*, 212–215. [CrossRef]
9. Benoît, C.; Mohamadi, T.; Kandel, S. Effects of Phonetic Context on Audio-Visual Intelligibility of French. *J. Speech Lang. Hear. Res.* **1994**, *37*, 1195–1203. [CrossRef]
10. Fort, M.; Spinelli, E.; Savariaux, C.; Kandel, S. The word superiority effect in audiovisual speech perception. *Speech Commun.* **2010**, *52*, 525–532. [CrossRef]
11. Fort, M.; Spinelli, E.; Savariaux, C.; Kandel, S. Audiovisual vowel monitoring and the word superiority effect in children. *Int. J. Behav. Dev.* **2012**, *36*, 457–467. [CrossRef]
12. Arnold, P.; Hill, F. Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *Br. J. Psychol.* **2001**, *92*, 339–355. [CrossRef]
13. Reisberg, D. Looking where you listen: Visual cues and auditory attention. *Acta Psychol.* **1978**, *42*, 331–341. [CrossRef] [PubMed]
14. Kawase, S.; Davis, C.; Kim, J. A Visual Speech Intelligibility Benefit Based on Speech Rhythm. *Brain Sci.* **2023**, *13*, 932. [CrossRef] [PubMed]
15. Vatikiotis-Bateson, E.; Eigsti, I.-M.; Yano, S.; Munhall, K.G. Eye movement of perceivers during audiovisualspeech perception. *Percept. Psychophys.* **1998**, *60*, 926–940. [CrossRef]
16. Hadley, L.V.; Brimijoin, W.O.; Whitmer, W.M. Speech, movement, and gaze behaviours during dyadic conversation in noise. *Sci. Rep.* **2019**, *9*, 10451. [CrossRef]
17. Lansing, C.R.; McConkie, G.W. Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Percept. Psychophys.* **2003**, *65*, 536–552. [CrossRef]
18. Barenholtz, E.; Mavica, L.; Lewkowicz, D.J. Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition* **2016**, *147*, 100–105. [CrossRef] [PubMed]
19. Birulés, J.; Bosch, L.; Pons, F.; Lewkowicz, D.J. Highly proficient L2 speakers still need to attend to a talker's mouth when processing L2 speech. *Lang. Cogn. Neurosci.* **2020**, *35*, 1314–1325. [CrossRef]
20. Grüter, T.; Kim, J.; Nishizawa, H.; Wang, J.; Alzahrani, R.; Chang, Y.-T.; Nguyen, H.; Nuesser, M.; Ohba, A.; Roos, S.; et al. Language proficiency modulates listeners' selective attention to a talker's mouth: A conceptual replication of Birulés et al. (2020). *Stud. Second. Lang. Acquis.* **2023**, *147*, 1–16. [CrossRef]
21. Lusk, L.G.; Mitchel, A.D. Differential Gaze Patterns on Eyes and Mouth During Audiovisual Speech Segmentation. *Front. Psychol.* **2016**, *7*, 52. [CrossRef]
22. Võ, M.L.-H.; Smith, T.J.; Mital, P.K.; Henderson, J.M. Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *J. Vis.* **2012**, *12*, 3. [CrossRef] [PubMed]
23. Buchan, J.N.; Paré, M.; Munhall, K.G. Spatial statistics of gaze fixations during dynamic face processing. *Soc. Neurosci.* **2007**, *2*, 1–13. [CrossRef] [PubMed]
24. Jayaraman, S.; Fausey, C.M.; Smith, L.B. The Faces in Infant-Perspective Scenes Change over the First Year of Life. *PLoS ONE* **2015**, *10*, e0123780. [CrossRef] [PubMed]
25. Jayaraman, S.; Smith, L.B. Faces in early visual environments are persistent not just frequent. *Vis. Res.* **2019**, *157*, 213–221. [CrossRef] [PubMed]
26. Morton, J.; Johnson, M.H. *Psychological Review CONSPEC and CONLERN: A Two-Process Theory of Infant Face Recognition*; Psychological Association, Inc.: Sydney, NSW, Australia, 1991.
27. Goren, C.C.; Sarty, M.; Wu, P.Y. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* **1975**, *56*, 544–549. [CrossRef] [PubMed]
28. Johnson, M.H.; Dziurawiec, S.; Ellis, H.; Morton, J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* **1991**, *40*, 1–19. [CrossRef]
29. TFarroni, T.; Johnson, M.H.; Menon, E.; Zulian, L.; Faraguna, D.; Csibra, G. Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17245–17250. [CrossRef]
30. Boothe, R.G.; Dobson, V.; Teller, D.Y. Postnatal Development of Vision in Human and Nonhuman Primates. *Annu. Rev. Neurosci.* **1985**, *8*, 495–545. [CrossRef]
31. Frank, M.C.; Vul, E.; Johnson, S.P. Development of infants' attention to faces duing the first year. *Cognition* **2009**, *110*, 160–170. [CrossRef]
32. Kwon, M.-K.; Setoodehnia, M.; Baek, J.; Luck, S.J.; Oakes, L.M. The development of visual search in infancy: Attention to faces versus salience. *Dev. Psychol.* **2016**, *52*, 537–555. [CrossRef]
33. Farroni, T.; Csibra, G.; Simion, F.; Johnson, M.H. Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9602–9605. [CrossRef] [PubMed]

34. Vecera, S.P.; Johnson, M.H. Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Vis. Cogn.* **1995**, *2*, 59–87. [CrossRef]
35. Hoehl, S.; Parise, E.; Palumbo, L.; Reid, V.M.; Handl, A.; Striano, T. Looking at Eye Gaze Processing and Its Neural Correlates in Infancy: Implications for Social Development and Autism Spectrum Disorder. *Socitety Res. Child Dev.* **2009**, *80*, 968–985. [CrossRef]
36. Elsabbagh, M.; Volein, A.; Csibra, G.; Holmboe, K.; Garwood, H.; Tucker, L.; Krljes, S.; Baron-Cohen, S.; Bolton, P.; Charman, T.; et al. Neural Correlates of Eye Gaze Processing in the Infant Broader Autism Phenotype. *Biol. Psychiatry* **2009**, *65*, 31–38. [CrossRef] [PubMed]
37. Mundy, P.; Newell, L. Attention, Joint Attention, and Social Cognition. *Curr. Dir. Psychol. Sci.* **2007**, *16*, 269–274. [CrossRef]
38. Senju, A.; Johnson, M.H. The eye contact effect: Mechanisms and development. *Trends Cogn. Sci.* **2009**, *13*, 127–134. [CrossRef]
39. Bastianello, T.; Keren-Portnoy, T.; Majorano, M.; Vihman, M. Infant looking preferences towards dynamic faces: A systematic review. *Infant Behav. Dev.* **2022**, *67*, 101709. [CrossRef]
40. Belteki, Z.; van den Boomen, C.; Junge, C. Face-to-face contact during infancy: How the development of gaze to faces feeds into infants' vocabulary outcomes. *Front. Psychol.* **2022**, *13*, 997186. [CrossRef]
41. Çetinçelik, M.; Rowland, C.F.; Snijders, T.M. Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development. *Front. Psychol.* **2021**, *11*, 589096. [CrossRef]
42. Lewkowicz, D.J.; Hansen-Tift, A.M. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1431–1436. [CrossRef]
43. Tenenbaum, E.J.; Sobel, D.M.; Sheinkopf, S.J.; Malle, B.F.; Morgan, J.L. Attention to the mouth and gaze following in infancy predict language development. *J. Child Lang.* **2015**, *42*, 1173–1190. [CrossRef]
44. ILozano, I.; Pérez, D.L.; Laudańska, Z.; Malinowska-Korczak, A.; Szmytke, M.; Radkowska, A.; Tomalski, P. Changes in selective attention to articulating mouth across infancy: Sex differences and associations with language outcomes. *Infancy* **2022**, *27*, 1132–1153. [CrossRef]
45. Tsang, T.; Atagi, N.; Johnson, S.P. Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *J. Exp. Child Psychol.* **2018**, *169*, 93–109. [CrossRef] [PubMed]
46. Morin-Lessard, E.; Poulin-Dubois, D.; Segalowitz, N.; Byers-Heinlein, K. Selective Attention to the Mouth of Talking Faces in Monolinguals and Bilinguals Aged 5 Months to 5 Years. *Dev. Psychol.* **2019**, *5*, 1640–1655. [CrossRef] [PubMed]
47. Pons, F.; Bosch, L.; Lewkowicz, D.J. Bilingualism Modulates Infants' Selective Attention to the Mouth of a Talking Face. *Psychol. Sci.* **2015**, *26*, 490–498. [CrossRef] [PubMed]
48. Imafuku, M.; Kanakogi, Y.; Butler, D.; Myowa, M. Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Dev. Sci.* **2019**, *22*, e12825. [CrossRef] [PubMed]
49. Farroni, T.; Menon, E.; Johnson, M.H. Factors influencing newborns' preference for faces with eye contact. *J. Exp. Child Psychol.* **2006**, *95*, 298–308. [CrossRef]
50. Senju, A.; Csibra, G. Gaze Following in Human Infants Depends on Communicative Signals. *Curr. Biol.* **2008**, *18*, 668–671. [CrossRef]
51. Ishikawa, M.; Senju, A.; Kato, M.; Itakura, S. Physiological arousal explains infant gaze following in various social contexts. *R. Soc. Open Sci.* **2022**, *9*, 35991332. [CrossRef]
52. Meltzoff, A.N.; Kuhl, P.K.; Movellan, J.; Sejnowski, T.J. Foundations for a New Science of Learning. *Science* **2009**, *325*, 284–288. [CrossRef]
53. Quine, W.V. Word and Object. *Philos. Phenomenol. Res.* **1960**, *22*, 115. [CrossRef]
54. Brooks, R.; Meltzoff, A.N. The development of gaze following and its relation to language. *Dev. Sci.* **2005**, *8*, 535–543. [CrossRef]
55. Brooks, R.; Meltzoff, A.N. Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *J. Child Lang.* **2008**, *35*, 207–220. [CrossRef] [PubMed]
56. Brooks, R.; Meltzoff, A.N. Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *J. Exp. Child Psychol.* **2014**, *130*, 67–78. [CrossRef]
57. Baldwin, D.A. Early Referential Understanding: Infants' Ability to Recognize Referential Acts for What They Are. *Dev. Psychol.* **1993**, *29*, 832–843. [CrossRef]
58. Munhall, K.; Johnson, E. Speech Perception: When to Put Your Money Where the Mouth Is. *Curr. Biol.* **2012**, *22*, R190–R192. [CrossRef]
59. Kuhl, P.K.; Meltzoff, A.N. The Bimodal Perception of Speech In Infancy. *Science* **1982**, *218*, 1138–1141. [CrossRef] [PubMed]
60. Patterson, M.L.; Werker, J.F. Two-month-old infants match phonetic information in lips and voice. *Dev. Sci.* **2003**, *6*, 191–196. [CrossRef]
61. Rosenblum, L.D.; Schmuckler, M.A.; Johnson, J.A. The McGurk effect in infants. *Percept. Psychophys.* **1997**, *59*, 347–357. [CrossRef] [PubMed]
62. Teinonen, T.; Aslin, R.N.; Alku, P.; Csibra, G. Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition* **2008**, *108*, 850–855. [CrossRef]
63. Ter Schure, S.; Junge, C.; Boersma, P. Discriminating Non-native Vowels on the Basis of Multimodal, Auditory or Visual Information: Effects on Infants' Looking Patterns and Discrimination. *Front. Psychol.* **2016**, *7*, 525. [CrossRef]

64. Tan, S.H.J.; Kalashnikova, M.; Burnham, D. Seeing a talking face matters: Infants' segmentation of continuous auditory-visual speech. *Infancy* **2022**, *28*, 277–300. [CrossRef] [PubMed]
65. Weatherhead, D.; White, K.S. Read my lips: Visual speech influences word processing in infants. *Cognition* **2017**, *160*, 103–109. [CrossRef] [PubMed]
66. Weatherhead, D.; Arredondo, M.M.; Garcia, L.N.; Werker, J.F. The Role of Audiovisual Speech in Fast-Mapping and Novel Word Retention in Monolingual and Bilingual 24-Month-Olds. *Brain Sci.* **2021**, *11*, 114. [CrossRef]
67. Haith, M.M.; Bergman, T.; Moore, M.J. Eye Contact and Face Scanning in Early Infancy. *Science* **1977**, *198*, 853–855. [CrossRef] [PubMed]
68. Colombo, J. The Development of Visual Attention in Infancy. *Annu. Rev. Psychol.* **2001**, *52*, 337–367. [CrossRef]
69. Amado, M.P.; Greenwood, E.; James; Labendzki, P.; Haresign, I.M.; Northrop, T.; Phillips, E.; Viswanathan, N.; Whitehorn, M.; Jones, E.J.H.; et al. The neural and physiological substrates of real-world attention change across development. *OSF Prepr.* **2023**. *preprint*. [CrossRef]
70. Alviar, C.; Sahoo, M.; Edwards, L.; Jones, W.R.; Klin, A.; Lense, M.D. Infant-directed song potentiates infants' selective attention to adults' mouths over the first year of life. *Dev. Sci.* **2023**. *online ahead of print*. [CrossRef]
71. Frank, M.C.; Vul, E.; Saxe, R. Measuring the Development of Social Attention Using Free-Viewing. *Infancy* **2012**, *17*, 355–375. [CrossRef]
72. Tenenbaum, E.J.; Shah, R.J.; Sobel, D.M.; Malle, B.F.; Morgan, J.L. Increased Focus on the Mouth Among Infants in the First Year of Life: A Longitudinal Eye-Tracking Study. *Infancy* **2012**, *18*, 534–553. [CrossRef]
73. Young, G.S.; Merin, N.; Rogers, S.J.; Ozonoff, S. Gaze Behaviour and Affect at 6 Months: Predicting Clinical Outcomes and Language Development in Typically Developing Infants and Infants At-Risk for Autism. *Dev. Sci.* **2009**, *12*, 798–814. [CrossRef] [PubMed]
74. Jones, W.; Klin, A. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* **2013**, *504*, 427–431. [CrossRef] [PubMed]
75. de Boisferon, A.H.; Tift, A.H.; Minar, N.J.; Lewkowicz, D.J. The redeployment of attention to the mouth of a talking face during the second year of life. *J. Exp. Child Psychol.* **2018**, *172*, 189–200. [CrossRef] [PubMed]
76. Birulés, J.; Bosch, L.; Brieke, R.; Pons, F.; Lewkowicz, D.J. Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Dev. Sci.* **2018**, *22*, e12755. [CrossRef] [PubMed]
77. Król, M.E. Auditory noise increases the allocation of attention to the mouth, and the eyes pay the price: An eye-tracking study. *PLoS ONE* **2018**, *13*, e0194491. [CrossRef]
78. Nakano, T.; Tanaka, K.; Endo, Y.; Yamane, Y.; Yamamoto, T.; Nakano, Y.; Ohta, H.; Kato, N.; Kitazawa, S.; Tamami, N.; et al. Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proc. R. Soc. B Boil. Sci.* **2010**, *277*, 2935–2943. [CrossRef]
79. Consortium ManyBabies. Quantifying sources of variability in infancy research using the infant-directed speech preference. *Adv. Methods Pract. Psychol. Sci.* **2020**, *3*, 24–52. [CrossRef]
80. Lense, M.D.; Shultz, S.; Astésano, C.; Jones, W. Music of infant-directed singing entrains infants' social visual behavior. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2116967119. [CrossRef]
81. Roth, K.C.; Clayton, K.R.H.; Reynolds, G.D. Infant selective attention to native and non-native audiovisual speech. *Sci. Rep.* **2022**, *12*, 15781. [CrossRef]
82. Birulés, J.; Martinez-Alvarez, A.; Lewkowicz, D.J.; de Diego-Balaguer, R.; Pons, F. Violation of non-adjacent rule dependencies elicits greater attention to a talker's mouth in 15-month-old infants. *Infancy* **2022**, *27*, 963–971. [CrossRef]
83. Kubicek, C.; de Boisferon, A.H.; Dupierrix, E.; Lœvenbruck, H.; Gervain, J.; Schwarzer, G. Face-scanning behavior to silently-talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. *Int. J. Behav. Dev.* **2013**, *37*, 106–110. [CrossRef]
84. Fort, M.; Ayneto-Gimeno, A.; Escrichs, A.; Sebastian-Galles, N. Impact of Bilingualism on Infants' Ability to Learn From Talking and Nontalking Faces. *Lang. Learn.* **2017**, *68*, 31–57. [CrossRef]
85. Ayneto, A.; Sebastian-Galles, N. The influence of bilingualism on the preference for the mouth region of dynamic faces. *Dev. Sci.* **2016**, *20*, 27196790. [CrossRef]
86. Pons, F.; Sanz-Torrent, M.; Ferinu, L.; Birulés, J.; Andreu, L. Children With SLI Can Exhibit Reduced Attention to a Talker's Mouth. *Lang. Learn.* **2018**, *68*, 180–192. [CrossRef]
87. SWass, S.V.; Goupil, L. Studying the Developing Brain in Real-World Contexts: Moving From Castles in the Air to Castles on the Ground. *Front. Integr. Neurosci.* **2022**, *16*, 896919. [CrossRef]
88. Bahrick, L.E.; Walker, A.S.; Neisser, U. Selective looking by infants. *Cogn. Psychol.* **1981**, *13*, 377–390. [CrossRef]
89. Bahrick, L.E.; Lickliter, R. Learning to Attend Selectively. *Curr. Dir. Psychol. Sci.* **2014**, *23*, 414–420. [CrossRef]
90. Alsius, A.; Navarra, J.; Campbell, R.; Soto-Faraco, S. Audiovisual Integration of Speech Falters under High Attention Demands. *Curr. Biol.* **2005**, *15*, 839–843. [CrossRef]
91. Tiippana, K.; Andersen, T.S.; Sams, M. Visual attention modulates audiovisual speech perception. *Eur. J. Cogn. Psychol.* **2004**, *16*, 457–472. [CrossRef]
92. Andersen, T.S.; Tiippana, K.; Laarni, J.; Kojo, I.; Sams, M. The role of visual spatial attention in audiovisual speech perception. *Speech Commun.* **2009**, *51*, 184–193. [CrossRef]

93. Lewkowicz, D.J.; Schmuckler, M.; Agrawal, V. The multisensory cocktail party problem in adults: Perceptual segregation of talking faces on the basis of audiovisual temporal synchrony. *Cognition* **2021**, *214*, 104743. [CrossRef]
94. Lewkowicz, D.J.; Schmuckler, M.; Agrawal, V. The multisensory cocktail party problem in children: Synchrony-based segregation of multiple talking faces improves in early childhood. *Cognition* **2022**, *228*, 105226. [CrossRef] [PubMed]
95. de Diego-Balaguer, R.; Martinez-Alvarez, A.; Pons, F. Temporal Attention as a Scaffold for Language Development. *Front. Psychol.* **2016**, *7*, 44. [CrossRef] [PubMed]
96. Gluckman, M.; Johnson, S.P. Attentional capture by social stimuli in young infants. *Front. Psychol.* **2013**, *4*, 527. [CrossRef] [PubMed]
97. Peltola, M.J.; Yrttiaho, S.; Leppänen, J.M. Infants' attention bias to faces as an early marker of social development. *Dev. Sci.* **2018**, *21*, e12687. [CrossRef]
98. Amso, D.; Haas, S.; Markant, J. An Eye Tracking Investigation of Developmental Change in Bottom-up Attention Orienting to Faces in Cluttered Natural Scenes. *PLoS ONE* **2014**, *9*, e85701. [CrossRef]
99. Frank, M.C.; Amso, D.; Johnson, S.P. Visual search and attention to faces during early infancy. *J. Exp. Child Psychol.* **2013**, *118*, 13–26. [CrossRef]
100. Chawarska, K.; Macari, S.; Shic, F. Decreased Spontaneous Attention to Social Scenes in 6-Month-Old Infants Later Diagnosed with Autism Spectrum Disorders. *Biol. Psychiatry* **2013**, *74*, 195–203. [CrossRef]
101. Abney, D.H.; Suanda, S.H.; Smith, L.B.; Yu, C. What are the building blocks of parent–infant coordinated attention in free-flowing interaction? *Infancy* **2020**, *25*, 871–887. [CrossRef]
102. Franchak, J.M.; Kretch, K.S.; Adolph, K.E. See and be seen: Infant–caregiver social looking during locomotor free play. *Dev. Sci.* **2017**, *21*, e12626. [CrossRef]
103. Yu, C.; Smith, L.B. Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE* **2013**, *8*, e79659. [CrossRef] [PubMed]
104. Clerkin, E.M.; Hart, E.; Rehg, J.M.; Yu, C.; Smith, L.B. Real-world visual statistics and infants' first-learned object names. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160055. [CrossRef] [PubMed]
105. Yurkovic-Harding, J.; Lisandrelli, G.; Shaffer, R.C.; Dominick, K.C.; Pedapati, E.V.; Erickson, C.A.; Yu, C.; Kennedy, D.P. Children with ASD establish joint attention during free-flowing toy play without face looks. *Curr. Biol.* **2022**, *32*, 2739–2746.e4. [CrossRef] [PubMed]
106. Deák, G.O.; Krasno, A.M.; Jasso, H.; Triesch, J. What Leads To Shared Attention? Maternal Cues and Infant Responses During Object Play. *Infancy* **2017**, *23*, 4–28. [CrossRef]
107. Fausey, C.M.; Jayaraman, S.; Smith, L.B. From faces to hands: Changing visual input in the first two years. *Cognition* **2016**, *152*, 101–107. [CrossRef]
108. Smith, L.B.; Jayaraman, S.; Clerkin, E.; Yu, C. The Developing Infant Creates a Curriculum for Statistical Learning. *Trends Cogn. Sci.* **2018**, *22*, 325–336. [CrossRef]
109. Schroer, S.E.; Yu, C. Looking is not enough: Multimodal attention supports the real-time learning of new words. *Dev. Sci.* **2022**, *26*, e13290. [CrossRef]
110. Mason, G.M.; Kirkpatrick, F.; Schwade, J.A.; Goldstein, M.H. The Role of Dyadic Coordination in Organizing Visual Attention in 5-Month-Old Infants. *Infancy* **2019**, *24*, 162–186. [CrossRef]
111. Goldstein, M.H.; Schwade, J.A. Social Feedback to Infants' Babbling Facilitates Rapid Phonological Learning. *Psychol. Sci.* **2008**, *19*, 515–523. [CrossRef]
112. Tamis-LeMonda, C.S.; Bornstein, M.H.; Baumwell, L. Maternal Responsiveness and Children's Achievement of Language Milestones. *Child Dev.* **2001**, *72*, 748–767. [CrossRef]
113. Tamis-LeMonda, C.S.; Kuchirko, Y.; Song, L. Why Is Infant Language Learning Facilitated by Parental Responsiveness? *Curr. Dir. Psychol. Sci.* **2014**, *23*, 121–126. [CrossRef]
114. Luchkina, E.; Xu, F. From social contingency to verbal reference: A constructivist hypothesis. *Psychol. Rev.* **2021**, *129*, 890–909. [CrossRef] [PubMed]
115. Yu, C.; Smith, L.B. The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Curr. Biol.* **2016**, *26*, 1235–1240. [CrossRef] [PubMed]
116. Wass, S.V.; Clackson, K.; Georgieva, S.D.; Brightman, L.; Nutbrown, R.; Leong, V. Infants' visual sustained attention is higher during joint play than solo play: Is this due to increased endogenous attention control or exogenous stimulus capture? *Dev. Sci.* **2018**, *21*, e12667. [CrossRef]
117. Wass, S.V.; Noreika, V.; Georgieva, S.; Clackson, K.; Brightman, L.; Nutbrown, R.; Covarrubias, L.S.; Leong, V. Parental neural responsivity to infants' visual attention: How mature brains influence immature brains during social interaction. *PLoS Biol.* **2018**, *16*, e2006328. [CrossRef]
118. Téglás, E.; Kovács, Á.M.; Gergely, G. Dissociating Measures of Information- and Control-Seeking in 12-Month-Olds' Contingency Exploration. In *Language, Cognition, and Mind*; Springer Nature: Berlin/Heidelberg, Germany, 2022; pp. 335–349. [CrossRef]
119. Yu, C.; Smith, L.B. Embodied attention and word learning by toddlers. *Cognition* **2012**, *125*, 244–262. [CrossRef]
120. Kuhl, P.K.; Tsao, F.-M.; Liu, H.-M. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9096–9101. [CrossRef] [PubMed]

121. Wass, S.V.; Whitehorn, M.; Haresign, I.M.; Phillips, E.; Leong, V. Interpersonal Neural Entrainment during Early Social Interaction. *Trends Cogn. Sci.* **2020**, *24*, 329–342. [CrossRef]
122. Roseberry, S.; Hirsh-Pasek, K.; Golinkoff, R.M. Skype Me! Socially Contingent Interactions Help Toddlers Learn Language. *Child Dev.* **2013**, *85*, 956–970. [CrossRef] [PubMed]
123. Vygotsky, L. *Mind and Society*; Harvard University Press: Cambridge, MA, USA, 1930.
124. Gratier, M.; Devouche, E.; Guellai, B.; Infanti, R.; Yilmaz, E.; Parlato-Oliveira, E. Early development of turn-taking in vocal interaction between mothers and infants. *Front. Psychol.* **2015**, *6*, 1167. [CrossRef] [PubMed]
125. Donnelly, S.; Kidd, E. The Longitudinal Relationship Between Conversational Turn-Taking and Vocabulary Growth in Early Language Development. *Child Dev.* **2021**, *92*, 609–625. [CrossRef] [PubMed]
126. Smith, N.A.; Trainor, L.J. Infant-Directed Speech Is Modulated by Infant Feedback. *Infancy* **2008**, *13*, 410–420. [CrossRef]
127. Kriengwatana, B.; Escudero, P.; Cate, C.T. Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization. *Front. Psychol.* **2015**, *5*, 1543. [CrossRef]
128. Polka, L.; Bohn, O.-S.; Weiss, D.J. Commentary: Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization. *Front. Psychol.* **2015**, *6*, 941. [CrossRef]
129. Pascalis, O.; Fort, M.; Quinn, P.C. Development of face processing: Are there critical or sensitive periods? *Curr. Opin. Behav. Sci.* **2020**, *36*, 7–12. [CrossRef]
130. Begus, K.; Gliga, T.; Southgate, V. Infants' preferences for native speakers are associated with an expectation of information. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12397–12402. [CrossRef]
131. Kaplan, P.S.; Jung, P.C.; Ryther, J.S.; Zarlengo-Strouse, P. Infant-Directed Versus Adult-Directed Speech as Signals for Faces. *Dev. Psychol.* **1996**, *32*, 880–891. [CrossRef]
132. Kaplan, P.S.; Zarlengo-Strouse, P.; Kirk, L.S.; Angel, C.L. Selective and Nonselective Associations Between Speech Segments and Faces in Human Infants. *Dev. Psychol.* **1997**, *33*, 990–999. [CrossRef]
133. Nencheva, M.L.; Lew-Williams, C. Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Dev. Rev.* **2022**, *66*, 101047. [CrossRef]
134. Oakes, L.M.; Plumert, J.M.; Lansink, J.M.; Merryman, J.D. Evidence for Task-Dependent Categorization in Infancy. *Infant Behav. Dev.* **1996**, *19*, 425–440. [CrossRef]
135. Madole, K.L.; Oakes, L.M.; Cohen, L.B. Developmental changes in infants' attention to function and form-function correlations. *Cogn. Dev.* **1993**, *8*, 189–209. [CrossRef]
136. Franchak, J.M.; Kretch, K.S.; Soska, K.C.; Adolph, K.E. Head-Mounted Eye Tracking: A New Method to Describe Infant Looking. *Child Dev.* **2011**, *82*, 1738–1750. [CrossRef] [PubMed]
137. LGoupil, L.; Proust, J. Curiosity as a metacognitive feeling. *Cognition* **2023**, *231*, 105325. [CrossRef] [PubMed]
138. Poli, F.; Serino, G.; Mars, R.B.; Hunnius, S. Infants tailor their attention to maximize learning. *Sci. Adv.* **2020**, *6*, eabb5053. [CrossRef] [PubMed]
139. Baer, C.; Kidd, C. Learning with certainty in childhood. *Trends Cogn. Sci.* **2022**, *26*, 887–896. [CrossRef] [PubMed]
140. Colomer, M.; Woodward, A. Should I learn from you? Seeing expectancy violations about action efficiency hinders social learning in infancy. *Cognition* **2023**, *230*, 105293. [CrossRef] [PubMed]
141. Feldman, R.; Masalha, S.; Alony, D. Microregulatory patterns of family interactions: Cultural pathways to toddlers' self-regulation. *J. Fam. Psychol.* **2006**, *20*, 614–623. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.