Article

FIRST LANGUAGE

First Language
2021, Vol. 41(2) 218–239

DOI: 10.1177/0142723720966815
journals.sagepub.com/home/fla
SAGE

Efficiency of scanning and attention to faces in infancy independently predict language development in a multiethnic and bilingual sample of 2-year-olds

David López Pérez*
Przemysław Tomalski*
Alicja Radkowska
Polish Academy of Sciences, Poland;
University of Warsaw, Poland

Haiko Ballieux
University of Westminster, UK

Derek G. Moore
University of Greenwich, UK

TALBY Team (Elena Kushnerenko, Deirdre Birtles, Mark H. Johnson and Annette Karmiloff-Smith)

---

*David López Pérez and Przemysław Tomalski have contributed equally to this work.

Corresponding authors:
David López Pérez, Institute of Psychology, Polish Academy of Sciences, Jaracza 1, 00-378 Warsaw, Poland.
Email: d.lopez@psych.pan.pl

Przemysław Tomalski, Institute of Psychology, Polish Academy of Sciences, Jaracza 1, 00-378 Warsaw, Poland.
Email: ptomalski@psych.pan.pl

Alicja Radkowska, University of Warsaw, Stawki 5/7, 00-183, Warsaw, Poland.
Email: a.radkowska@psych.uw.edu.pl

Abstract

Efficient visual exploration in infancy is essential for cognitive and language development. It allows infants to participate in social interactions by attending to faces and learning about objects of interest. Visual scanning of scenes depends on a number of factors, and early differences in efficiency are likely contributing to differences in learning and language development during subsequent years. Predicting language development in diverse samples is particularly challenging, as additional multiple sources of variability affect infant performance. In this study, we tested how the complexity of visual scanning in the presence or absence of a face at 6 to 7 months of age is related to language development at 2 years of age in a multiethnic and predominantly bilingual sample from diverse socioeconomic backgrounds. We used Recurrence Quantification Analysis to measure the temporal and spatial distribution of fixations recurring in the same area of a visual scene. We found that in the absence of a face the temporal distribution of re-fixations on selected objects of interest (but not all) significantly predicted both receptive and expressive language scores, explaining 16% to 20% of the variance. Also, lower rate of re-fixations recurring in the presence of a face predicted higher receptive language scores, suggesting larger vocabulary in infants that effectively disengage from faces. Altogether, our results suggest that dynamic measures, which quantify the complexity of visual scanning, can reliably and robustly predict language development in highly diverse samples. They suggest that selective attending to objects predicts language independently of attention to faces. As eye-tracking and language assessments were carried out in early intervention centres, our study demonstrates the utility of mobile eye-tracking setups for early detection of risk in attention and language development.

Keywords

Eye-Tracking, Recurrence Quantification Analysis, Visual Scanning, Socio-Economic Status, Language Development

Introduction

From early on infants constantly scan their visual environment in search for new information and objects to explore (Franchak et al., 2010; Yoshida & Smith, 2008). Visual scanning also helps infants to detect other humans in the vicinity, so that they can orient and establish social interactions (Johnson et al., 2015). Thus, effective visual exploration serves multiple purposes and plays a central role in human learning and development (Sokolov, 1963). Identifying how early differences in exploration lead to later outcomes may provide insight into different mechanistic pathways in cognitive development. Here, we investigated how two aspects of visual scanning in infancy may differentially predict receptive and expressive language in a multiethnic and predominantly bilingual sample of 2-year-olds.

First, preferential attention to faces and eye gaze facilitates the establishment of social interactions and as a consequence supports the development of language and communication. Newborns preferentially orient to faces and direct gaze (Johnson, 1991), and throughout the subsequent months they show increased attention to faces in complex scenes (Frank et al., 2009, 2014). With age, infants show increased sensitivity to social signals in faces, such as eye contact (e.g. Gredebäck et al., 2010; Hains & Muir, 1996; Parise et al., 2008) and use them effectively to build their attention skills (e.g. Niedźwiecka et al., 2017), communicative and language skills (Beier & Spelke, 2012; Brooks & Meltzoff, 2005; Parise et al., 2011) or learn about objects (Senju et al., 2008). Attention to faces supports language development in multiple ways – apart from facilitating social interactions and learning, it provides correlated multisensory experience necessary to develop phonological knowledge and vocabulary repertoire (e.g. (Kushnerenko et al., 2013; Tenenbaum et al., 2014). Thus, greater attention to faces in infancy is thought to predict better language outcomes later in life (e.g. Brooks & Meltzoff, 2005; Morales et al., 1998). However, in typical development, with age children learn to disengage from faces to look at other relevant aspects of social situations, such as hands (e.g. Frank et al., 2011; Yu & Smith, 2017). Studies of atypical development also suggest that inefficient attention shifting or disengagement may lead to longer looking at faces, thus likely disrupting the flow of social interactions and learning sequences (Elsabbagh et al., 2013; Hendry et al., 2018; Jones et al., 2018; Parsons et al., 2019; Wagner et al., 2016). These findings suggest that looking to faces is important for language development, but that it is also important to know when to disengage.

Second, attention development in infancy involves increasing efficiency of visual search, which is driven by curiosity, ambiguity or novelty (Perone & Spencer, 2013). Infants scan their visual environment in search of novel objects and events, and their novelty preference is manifested by longer looking (e.g. Bornstein, 1985), a phenomenon used widely in visual preference methods for studying their perception and conceptual knowledge. A novelty preference in visual habituation tasks has also been found to predict productive vocabulary (Colombo et al., 2004) and vocabulary growth at later ages (Marino & Gervain, 2019). However, individual differences in overall duration of looking may not reflect differences in the dynamics of scanning in more complex visual scenes with multiple objects (Anderson et al., 2013). Finding these individual differences between learners is important since it may shed more light on underlying learning mechanisms (e.g. Marino & Gervain, 2019; Newman et al., 2006). Therefore, in our study we employed dynamic measures that can help to capture individual differences in more complex patterns of visual scanning.

Altogether, it remains unclear to what extent infant visual scanning in the middle of the first year of life is driven primarily by exogenous vs endogenous mechanisms and to what extent the visual skills that rely on these mechanisms independently contribute to language development. On one hand, exogenously driven sensitivity to the presence of social stimuli predicts communicative skills, but the direction of this association is unclear. In our study, we treat attention to faces as exogenously driven, since infants direct their first saccades to faces presented among other objects as well as scanning them more extensively (higher number of fixations in face area) (Gliga et al., 2009). On the other hand, efficient novelty detection and visual processing of visual scenes, driven by internal goals, and lower distractibility (Salley et al., 2013) also contribute to language learning. Endogenously driven orienting involves voluntary or strategic gaze shifts. Four- to five-month-olds show voluntary control of eye movements and scan their environment more flexibly than younger infants (Hunnius, 2006). Therefore, we will treat attention to non-face objects in our task as predominantly endogenously driven.

Since previous literature has often focused on singular aspects of attention (e.g. fixating on faces), we investigated the relative contribution of both kinds of mechanisms in a single task, by using arrays of objects either containing a face or not. Moreover, instead of using traditional cumulative measures of looking, we quantified the individual differences in the complexity of scanning in both conditions within an individual to assess the relative contribution of different attentional skills to language development: the attention-getting and attention-holding by faces and the efficiency of visual search in the absence of a face.

To quantify the complexity of scanning, we used the newly developed dynamical measures (i.e. Recurrence Quantification Analysis; RQA) that account for the temporal and spatial distribution of repetitions of fixations (i.e. re-fixations) in the same area of an image (Anderson et al., 2013). The advantage of RQA in comparison to other methods is that it captures individual differences in temporal dynamics, to provide an in-depth depiction of what drives infants' attention, and to explore how these processes evolve over viewing time, which is something that cannot be investigated with traditional high-level measures (López Pérez et al., 2018). We measured attention to faces and efficiency of visual scanning by presenting infants with arrays of objects belonging to different categories that could either contain a static face ('face slides') or control arrays in which the face was replaced by a chair ('chair slides'). Previous analysis demonstrated the reliability of these measures and showed much higher trial-to-trial variability of scanning in the absence of a face by 6- to 7 month-old infants (López Pérez et al., 2018). Building on this work, we investigated (1) how often fixations recurred in the same area when a face was present (RR, Recurrence Rate – Faces); and (2) how fixations and re-fixations are temporally distributed throughout the trial, when a face was absent, measuring whether infants revisited some of the selected objects later in the trial (CORM, Centre of Recurrence Mass – Chairs). Given the exogenous orienting effect of faces (Gliga et al., 2009), and that attention to faces predicts language development, we expected the RR would be associated with both receptive and productive language scores, although the direction of this effect could not be specified on the basis of previous research. Likewise, in the absence of a face, we predicted higher variability in the infants' visual exploration strategies (López Pérez et al., 2018). In this case, the RR is less interesting given that there is no supporting evidence that any of the objects are particularly relevant for communicative skills. However, the CORM indicates differences in the temporal patterns of exploration and describes whether infants scan rapidly the entire scene followed by the selection and detailed scanning of objects of interest. In this case, we expected higher CORM to be related to higher language scores. We did not have differential predictions in relation to receptive vs productive language scores.

The question of language predictors is especially complex in the case of infants from bi- or multilingual environments. Their early language experience is much more diverse in terms of input, but also likely more variable between infants in comparison with monolingual infants due to varied contexts and persons that use specific languages. Also, from the middle of the first year of life, infants are likely to develop different strategies for allocating attentional resources based on their language environment (e.g. Comishen et al., 2019; Pons et al., 2015). This means that predicting language development of infants from bi- and multilingual families could be more challenging due to the significantly greater variability in their experience and attentional strategies (e.g. Place & Hoff, 2011). There is very limited data on the relation between early attention and later language development in such highly varied samples.

Here we tested a unique and highly diverse sample of infants in terms of the socioeconomic background and ethnicity, where nearly 75% of participating families were bi- or multilingual. During the visit, infants were given a visual scanning task at 6 months of age (T1) and then around 18 months later (T2) a subset of these children were administered language tasks assessing receptive and expressive language skills. We investigated whether new dynamic measures of scanning could help to resolve some of the outstanding problems with predicting communicative development in such samples, because such measures capture complex patterns of scanning, and are less sensitive to absolute differences in looking times or variability in individual fixation durations (López Pérez et al., 2018).

Methods

Participants

The eye-tracking assessments were conducted in community settings, in seven 'Sure-Start' Children's Centres (CCs) in East London (United Kingdom), located in two urban boroughs (Newham and Tower Hamlets) with some of the highest levels of deprivation nationwide (according to the English Index of Multiple Deprivation; Department for Communities and Local Government, 2010). Participants were recruited to take part in 'Learn about your Baby' sessions, which were part of the scheduled timetable of activities of the CCs (for more details on the sample and the study design, see Ballieux et al., 2016).

One hundred eighty-three infants were recruited to the study and their family socioeconomic status (SES) represented the population of this London area. Nine participants out of 183 originally recruited were subsequently excluded from the sample when researchers rechecked eligibility. Participants had a wide range of income and education levels, from very low levels of education and income to highly educated and affluent parents. Of the remaining 174 participants, 65 were rejected because they did not produce enough eye-tracking (ET) data (for inclusion criteria, see 'Data Pre-processing' below). The final sample analysed for the ET task at T1 consisted of 109 infants (sample descriptives are presented in Table 1). None of the participants had older siblings with autism or any major delivery complications or major medical conditions (genetic, metabolic, or other chronic illness). No mother reported using recreational drugs throughout pregnancy, while two reported smoking and 16 reported low levels of alcohol consumption (weekly level, range: 0.5–2 UK units). The study received ethical approval from the local university board and from local government authority and complied with the Declaration of Helsinki. All parents gave written informed consent and received small gifts in return for their participation.

We compared the T1 participants included ($n = 104$) and excluded ($n = 70$) because of the ET data quality on sociodemographics. There were no significant differences between the groups on family income ($t = .10$, $p = .92$), infant birthweight and gestational age (both ts $< .97$, ps $> .33$). They also did not differ in terms of maternal ($\chi^2 = .99$, $p = .60$) or

Table 1. Full demographics of the participants in the T1 sample, the T2 sample and the final longitudinal sample.

	Sample T1	Sample T2	Longitudinal analysis	
Participants (*n*)	109	83	45	
Age (days)			T1	T2
M	207.91	750.69	205.4	757.6
SD	21.59	44.36	20.1	48.4
Range	181–240	675–922	152–247	675–922
Gender				
Girls	40 (36.7%)	39 (47%)	18 (40%)	
Boys	69 (63.3%)	44 (53%)	27 (60%)	
Ethnicity				
Caucasian	25 (25.9%)	24 (28.9%)	11 (24.5%)	
Afro-Caribbean	14 (11.5%)	8 (9.6%)	5 (11.1%)	
Asian-Indian	45 (34.5%)	35 (42.2%)	22 (48.9%)	
Mixed ethnicity	25 (28.2%)	16 (19.3%)	7 (15.5%)	
Gestational age				
M	39.3	39.5	39.4	
SD	1.8	1.5	1.6	
Range	32–42	36–42	36–42	
Birthweight (g)				
M	2919.0	3020.1	3151.8	
SD	971.9	982.4	478.2	
Range	2250–4140	2250–5400	2250–3941	
Maternal age at birth (year)				
M	30.4	30.2	30.7	
SD	4.9	4.9	4.3	
Range	18–45	19–45	22–39	
Family language environment				
Monolingual	29 (26.6%)	24 (28.9%)	12 (26.7%)	
Bilingual or multilingual	80 (73.4%)	59 (71.1%)	33 (73.3%)	
English first language at home (*n*)	62 (56.9%)	45 (54.2%)	20 (44.4%)	

paternal social class ($\chi^2 = 3.32, p = .19$) and paternal education ($\chi^2 = .55, p = .46$), while in the included group there were marginally more mothers without high school education ($\chi^2 = 3.05, p = .081$). Finally, there were no significant differences between these groups in terms of maternal ethnicity ($\chi^2 = 2.26, p = .52$) or bilingualism ($\chi^2 = .96, p = .33$).

Follow-up data collection at T2. A total of 83 participants who were tested at T1 (24% participant loss) took part in the follow-up study when children were aged around 24 months (see Table 1). The second visit took place in the same CC as at T1, but due to changes in funding and re-organisation of Centres and their services, we were unable to conduct the follow-up in one of them. As certain free-of-charge services were no longer available to many low-income parents, Centre staff were no longer available to assist with contacting some parents and several caregivers moved out of the borough, we experienced a considerable dropout of study participants. Consequently, follow-up data were not available in 38 of the 83 participants tested at T2 (approximately 45% participant loss). However, the T2-tested group did not differ from the follow-up absent group in terms of socioeconomic indicators: gross family income (t=.49, p=.63), maternal education (χ^2=1.10, p=.30), paternal education (χ^2=.86, p=.35), and maternal (χ^2=.72, p=.70) or paternal social class (χ^2=.46, p=.79). There were no differences in terms of perinatal risk factors (gestational age and birthweight, both ts < 1.04, ps > .30) or in the maternal ethnicity (χ^2=.54, p=.91) or bilingualism (χ^2=.19, p=.66).

Final longitudinal sample. The final sample analysed longitudinally included 45 children (18 girls and 27 boys) tested as infants around the age of 6 to 7 months and followed-up again around the age of 24 months. All children were delivered at term and nearly all had birthweight within the normal range. Maternal age at birth was on average 30.7 years. The sample was uniquely multiethnic and was highly varied in terms of languages spoken at home, with the majority of parents reporting bilingual or multilingual family environment. For 20 children (44.4%), English was their first language at home. The sample also represented a wide range of socioeconomic and educational backgrounds.

We also compared language scores at T2 between the infants included and excluded due to ET data quality at T1 and did not find any differences for either Preschool Language Scale auditory comprehension (PLS AC) (t=.52, p=.61) or Preschool Language Scale expressive communication (PLC EC) scores (t=.052, p=.96). Data for expressive language measures in PLS were not available for 6 children out of 45 due to their inability to complete the assessment. Thus, for PLC EC we analysed n=39 and for PLC AC we analysed n=45 children.

Eye-tracking task and stimuli at T1

We used a modified (Ballieux et al., 2016; López Pérez et al., 2018) face pop-out task (Gliga et al., 2009), where infants freely viewed visual scenes of six coloured objects on a white background. Ten visual scenes were created, each of them containing six objects from different categories. Five objects, common among the 10 scenes, consisted of examples from categories of shoes, cars, mobiles, birds, and clocks. The remaining object was selected from two categories of objects: faces or chairs. Five visual scenes contained examples from the category of chairs, while the other five from the category of faces (four female and one male). All the faces displayed neutral expressions and the task was adapted for use with a diverse population including a wider variety of ethnicities of faces (Ballieux et al., 2016). Each scene was presented on the screen for 10 seconds. There were two different pseudorandom orders of presentation, where the 10 scenes were presented in two blocks, with the block order counterbalanced between subjects.

Eye-tracking procedure at T1

At T1, ET data were acquired using a portable kit, which contained a 17" eye-tracker with integrated monitor (Tobii T120) and a portable Ergotron MX desk mount arm that could be clamped onto a table and adjusted to provide consistency in the height of the screen relative to the position of the infant. An HP EliteBook 8440p laptop was used to control the eye-tracker using Tobii Studio version 2.0. The distance of the infant's head to the screen was 60 cm and the approximate height of the infant was 1.3 m (for further description of the protocol, see Ballieux et al., 2016).

Follow-up language measures at T2

The assessment of bi- and multilingual children poses many challenges, and a reliable estimate of their total vocabulary size would require the measurement for each language used in the family. Since this was not possible in the sample with so many languages, we opted for a direct, performance-based measure of language communicative skills, the Preschool Language Scale-4 (PLS-4; Zimmerman et al., 2002), which is a norm-referenced test of receptive and expressive language ability for ages from birth to 6 years. The test consists of a picture book and toys designed to engage a child to elicit responses to test items. It gives two standardised subscales, auditory comprehension (PLS AC) and expressive communication (PLS EC), and a total score. PLS-4 instructions are in English, but the manual explicitly allows for the caregiver to translate instructions for each item during the assessment into another language to ensure optimal child performance. As normative data for bilinguals were not available, we opted for using raw scores. See the descriptive data for language assessment in Table 2. Our secondary language outcome measure was Mac Arthur-Bates Communicative Development Inventory (CDI). Longitudinal associations closely resembled the results obtained with PLS-4 (see Supporting Information for description of these results).

Upon follow-up, the children were tested in the same room in each CC as during the first visit. The session consisted of one short eye-tracking task (not reported here), a joint attention standardised task (Early Social Communication Scales, not reported here) and language assessment using the PLS-4. Each session was videotaped for later verification of scoring.

Eye-tracking data analysis

Pre-processing

Trials were included in the analysis if at least 50% of the gaze samples for both eyes were valid and included at least five fixations, which is needed to be able to quantify some dynamics. These values allow infants to develop sufficient visual exploration behaviour that can be later related to their exploration strategies. Using a 50% threshold for valid trials guarantees that the infant is attentive to the screen and looking at it for a sufficient amount of time. Having a minimum of five fixations increases the possibility of having dynamic patterns of fixations, while saving most of the trials. In addition, participants that did not provide at least three valid trials for each visual scene type were rejected from the analysis sample. Having at least three valid trials ensures that the dynamic measures of visual exploration are a real feature of each type of scene and not

Table 2. Descriptive data for language measures at 24 months (T2).

	n	*M*	*SD*	Min	Max
PLS AC	45	26,91	3,71	19	34
PLS EC	39	27,82	3,15	20	33

Table 3. Hierarchical regression models of visual scanning dynamics in the pop-out task at 6 to 7 months of age predicting receptive and expressive language scores at 24 months of age in Preschool Language Scales-4.

Predictor	PLS Auditory Comprehension					PLS Expressive Communication				
	B	*SE*	Beta	*B* CI 95%		*B*	*SE*	Beta	*B* CI 95%	
				Lower limit	Upper limit				Lower limit	Upper limit
Step 1										
Constant	27.86	1.21		26.76	28.96	28.23	0.52		27.23	29.33
Income	3.96	1.21	0.46**	1.52	6.40	2.36	1.22	0.31†	−0.12	4.84
R^2	0.21**					0.1†				
Step 2										
Constant	30.30	1.30		27.68	32.92	28.02	1.39		25.20	30.85
Income	3.67	1.17	0.42**	1.30	6.04	2.37	1.24	0.31†	−0.15	4.88
RR Faces	−0.10	0.05	−0.28*	−0.19	−0.01	0.01	0.05	0.03	−0.1	−0.12
ΔR^2	0.08*					0.01				
R^2	0.28**					0.1				
Step 2										
Constant	27.70	1.39		24.88	30.52	25.81	1.43		22.90	28.71
Income	3.49	1.05	0.40**	1.37	5.61	2.55	1.11	0.34*	0.30	4.80
RR Faces	−0.14	0.04	−0.41**	−0.23	−0.05	−0.04	0.05	−0.12	−0.14	0.06
CORM Chairs	0.61	0.18	0.42**	0.24	0.98	0.57	0.18	0.48**	0.20	0.94
ΔR^2	0.16**					0.21**				
R^2	0.44***					0.30**				

$^{\dagger}p < .1$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

the consequence of one unique measurement (e.g. if only one trial was valid for one of the types of scenes). From the final sample of 109 participants, the average number of trials with faces was 4.35 (*SD* = 0.75, range 3–5) and for trials with chairs, it was 4.10 (*SD* = 0.86, range 3–5).

Prior to the RQA analysis, fixation coordinates and durations were extracted using a novel noise-robust fixation detection algorithm that uses 2-means clustering (Hessels et al., 2016). This algorithm can detect fixations in noisy data, which makes it suitable for infant research where data quality is generally poorer than adult studies (Hessels et al., 2015), and especially suitable for our data set, which was collected in community settings (Ballieux et al., 2016). We used most of the suggested default settings for the algorithm (see Hessels et al., 2016). For the Steffen interpolation, we used an interpolation window of 100 ms and an interpolation edge of 2 samples (i.e. 16.66 ms). We chose these values since values longer than 100 ms would lead to interpolation of blinks, which usually take longer than 100 ms (Hessels et al., 2016), while smaller values lead to less periods of data loss being interpolated. In the k-means clustering, we applied a sample-by-sample analysis, a clustering window size of 200 ms, downsampling to assure that the transitions between fixations are not caused by high-frequency noise in the data at 60, 30 and 15 Hz and a clustering cutoff of 2 times the standard deviation above the k-means weights. Given that fixation durations are typically longer than 150 ms (e.g. Irwin, 1992), the clustering window would contain parts of at most two fixations. Next, all those fixations that had a minimum duration of 100 ms were considered valid and shorter fixations candidates were excluded. We chose this conservative minimum fixation duration since longer values would lead to short fixation candidates being excluded and some studies have argued that fixations durations of 100 ms can also be justified (Manor & Gordon, 2003). Finally, we merged fixation candidates that were less than 0.7 degrees apart and separated by less than 30 ms. Increasing both parameters would lead to more fixations being merged.

Recurrence quantification analysis

The dynamics of visual scanning were explored using RQA on fixation sequences in pre-processed (fixation-filtered) ET data, an analysis developed to characterise the gaze patterns of single observers (Anderson et al., 2013). RQA describes the local and global properties of fixation sequences extracting a handful of parameters, which are sensitive to the type of scene and have a clear interpretation in the context in which they are extracted (Anderson et al., 2013; Wu et al., 2014). In a previous study, we introduced RQA to study visual exploration in infants and how the dynamics varied during a face pop-out task (for a detailed description of the RQA analysis, see López Pérez et al., 2018).

In this article, we primarily focused on the analysis of the global properties of fixations and its relation to later language developmental outcomes. In RQA, two fixations are considered to be recurrent if they are within a certain distance or radius of each other (see red circle in Figure 1(a)). Using this information, we can reconstruct the recurrence plot (see Figure 1(b)), which is a visual representation of all the recurrences of a fixation sequence with itself at all possible time lags. For instance, fixation 3 is recurrent with fixations 10 and 14 (see blue circle in Figure 1(b)), while at the same time fixation 14 is recurrent with fixation 10 (see green circle in Figure 1(b)). The local and global properties of fixation sequences can then be extracted by quantifying the different structures that arise in the recurrence plot (Figure 2). In particular, we chose the RR and the CORM because they quantify the overall visual exploration of a stimulus. The RR represents the percentage of fixations that are part of areas previously fixated or how often infants re-fixate previously fixated image positions (i.e. all red dots in the recurrence plot). A high RR value represents that fixations fall mostly in the same areas (see Figure 2(b)), while a low RR value is related to more dispersed fixation patterns (see Figure 2(d)). The larger the distance between the main diagonal and the recurrent fixation, the larger the time interval (in fixations) between the original

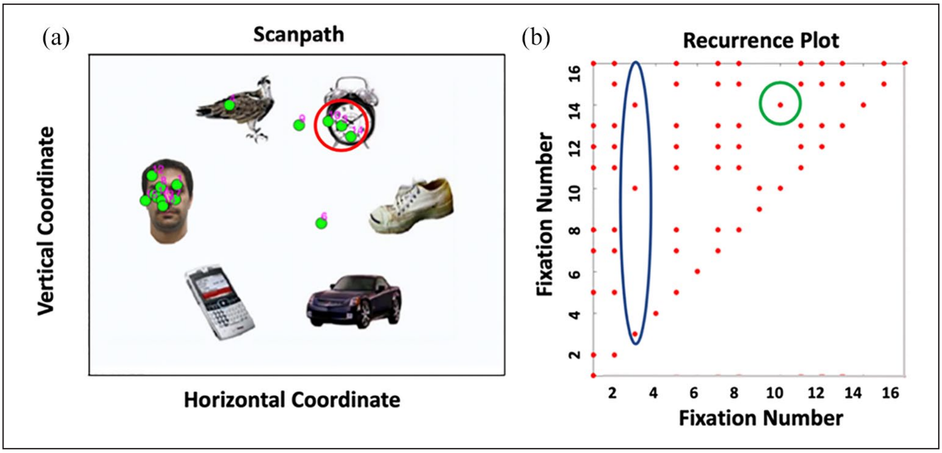

Figure 1. Example of a Fixation Scanpath in Face Slides. Most of the fixations were located on the face (a). The numbers in the scanpaths indicate the fixation order and each red dot in the recurrence plot (b) indicates a re-fixation in a previously fixated location (see fixations inside the red circle in a). The blue circle indicates how fixation 3 is recurrent with fixations 10 and 14 while the green one how at the same time fixation 14 is recurrent with fixation 10. The recurrence plot is symmetric and only the upper triangle is displayed. The quantitative measures are usually extracted excluding the line of incidence, which does not add any additional information since it indicates that each fixation is recurrent with itself. Written consent was given by the first author of this publication to use the image of his face for this figure.

fixation and subsequent re-fixations (Anderson et al., 2013). This temporal distribution of re-fixations is what the CORM quantifies, where low values indicate that re-fixations tend to occur close in time (see circle B in Figure 2(b) and (d)), whereas large values indicate that re-fixations tend to occur at longer intervals in time (see circle A in Figure 2(b) and (d)). Formally, CORM is defined as the distance of the centre of gravity of recurrent points from the line of incidence, normalised such that the maximum possible value is 100 (Anderson et al., 2013) and it is computed as follows

$$\mathrm{CORM}=100\frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(j-i)r_{ij}}{(N-1)R}$$

where j and i are the fixation positions in the sequence, r represents whether a given fixation was recurrent or not, N is the total number of fixations and R is the total number of recurrences. For instance, if an infant scans an area (e.g. the clock) in detail and returns to it later in the trial the $(j - i)$ will increase and the CORM will be higher because the recurrence points will be separated in time (see circle A in Figure 2(b)). Otherwise, if an infant scans particular areas of the scene and never returns to them later in the trial, most of the recurrent points will fall close to the line of incidence (see circle B in Figure 2(d)) and the $(j - i)$ will be small, which will be represented by a low CORM value. We

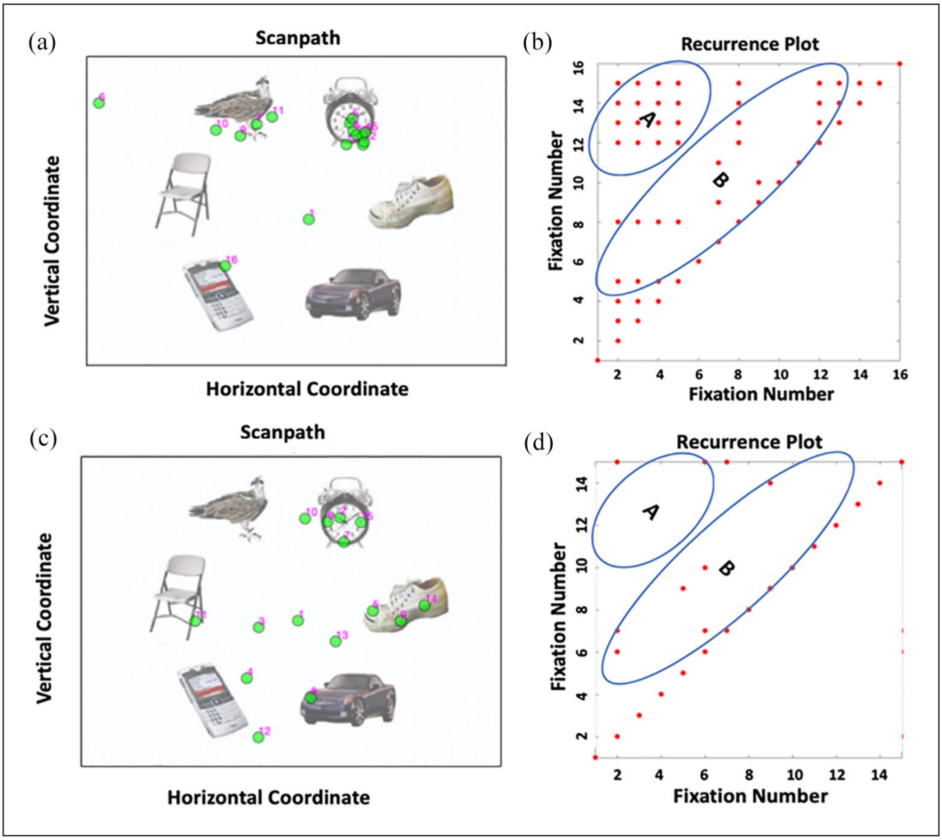

Figure 2. Examples of High (a and b) and Low (c and d) CORM Infants in Chair Slides. High CORM infants were characterised by high number of revisitations to previously fixated areas while in low CORM infants these number of revisitations and the temporal gaps between them were much smaller (differences between A and B). The numbers in the scanpaths indicate the fixation order and each red dot in the recurrence plot indicates a re-fixation in a previously fixated location. The quantitative measures are usually extracted excluding the line of incidence, which does not add any additional information since it indicates that each fixation is recurrent with itself.

focused on the RR on face slides and the CORM in chair slides. First, we chose for our model the RR measure for face slides because it is informative about preferential attention to faces (see example in Figure 1). However, since the exploration in face slides is exogenously driven by faces, it is less likely that the CORM will pick up significant differences in the temporal patterns of exploration (see López Pérez et al., 2018). Second, we focused on the CORM in chair slides because in the absence of a face, we were more interested in the visual exploration strategies and therefore in the global distribution of fixations (i.e. how fixations and re-fixations were temporally distributed). In this type of display, the RR would be less informative because it would only quantify which object is more interesting for each infant, while none of the objects are more relevant to language scores than others. Finally, to account for the individual variability in fixation durations, we normalised RQA measures with average fixation durations for each infant (Anderson et al., 2013).

Results

No effects of bilingualism, maternal ethnicity and socioeconomic status on eye-tracking predictors

In our final sample, there were no differences in RR for face slides or CORM for chair slides between bilingual and monolingual families (both ts < 0.63 ps $> .53$). Likewise, neither eye-tracking measure differed significantly depending on maternal ethnicity (both Fs < 1.6, ps $> .2$).

Analyses of socioeconomic indicators showed that there were no group differences in either predictor as a function of material social class (both Fs < 1.1, ps $> .37$), paternal social class (both Fs < 1.39, ps $> .26$), maternal education (both Fs < 0.13, ps $> .72$), or paternal education (both Fs < 0.16, ps $> .69$). Likewise, there were no significant correlations with family income (both rs $< .12$, ps $> .44$).

Dynamics of scanning differentially predict receptive vs productive language

We conducted hierarchical regression analyses separately for receptive and expressive language scores of PLS-4 (see Table S1 in Supporting Results for zero-order correlations). For both dependent variables, we tested the same model with gross family income, recurrence rate for face slides (RR Faces) and centre of recurrence mass for chair slides (CORM Chairs) entered in subsequent steps (Table 3).

For receptive language scores, all three variables at 6 to 7 months of age predicted a unique proportion of its variance. In the first step, family income ($\beta = .46$, $t = 3.27$, $p = .002$) significantly predicted nearly 21% the variance in receptive language scores, $R^2 = .207$, $F(1,41) = 10.71$, $p = .002$. In the second step, lower RR for face slides ($\beta = -.28$, $t = -2.06$, $p = .046$) predicted higher receptive scores, explaining an additional 7.6% of the variance, $\Delta R^2 = .076$, $F(1,40) = 4.24$, $p = .046$. Finally, in the last step, higher CORM for chair slides predicted higher receptive language ($\beta = .42$, $t = 3.36$, $p = .002$), explaining a further 16% of the variance, $\Delta R^2 = .161$, $F(1,39) = 11.27$, $p = .002$. Altogether, the entire model explained nearly 45% of the variance in receptive language raw scores with all three variables showing significant effects, $R^2 = .444$, Adj. $R^2 = .401$, $F(3,39) = 10.38$, $p < .001$.

For expressive language, the same model with three predictor variables was tested, but returned a different set of results. In the first step, there was a trend for higher family income ($\beta = .31$, $t = 1.93$, $p = .061$) to predict higher language scores, $R^2 = .096$, $F(1,35) = 3.74$, $p = .061$. In the second step, attention to faces (RR for face slides) did not significantly predict expressive language, $\beta = .03$, $t = 0.20$, $p = .84$; $\Delta R^2 = .001$, $F(1,34) = 0.04$, $p = .84$. Finally, in the last step, higher CORM for chair slides significantly predicted expressive language ($\beta = .48$, $t = 3.12$, $p = .004$), explaining nearly 21% of its variance, $\Delta R^2 = .206$, $F(1,33) = 9.75$, $p = .004$. The final model significantly explained nearly 30% of the variance in expressive language scores, $R^2 = .303$, Adj. $R^2 = .240$, $F(3,36) = 4.79$, $p = .007$.

Controlling for bilingualism. We further tested for the effects of bilingual (or multilingual) language environment on our regression models, by adding this variable as a predictor in the first step (see Supporting Results for details). Bilingualism status did not significantly predict either receptive or expressive language, confirming that PLS provides a less English language-biased assessment of communication in multilingual samples. The inclusion of bilingualism status in the first step of the regression analyses did not alter the pattern of results with respect to measures of visual scanning. After controlling for bilingualism, receptive language (PLS AC) was significantly predicted by RR for face slides and CORM for chair slides, while expressive language (PLS EC) was significantly predicted by CORM for chair slides alone, with a similar proportion of the variance being explained as in the models that did not control for bilingualism.

Secondary outcome measure. Additional regression analyses were performed for our English-biased measure of language outcomes – the Communicative Development Inventory. This measure was used as secondary, because it is less suitable for the assessment of communicative development in bilingual and multilingual children, for whom vocabulary assessment should involve all languages used at home. However, we used the proportion of time that English was spoken as an additional control measure to account for differences in language environment. The regression analyses confirmed the general pattern of results, where RR Faces and CORM Chairs independently predicted language scores at 2 years of age.

Discussion

Our results show that dynamic measures of infants' visual scanning significantly and robustly predict language outcomes a year and a half later in a linguistically and ethnically diverse group of toddlers. Lower Recurrence Rate of fixations in the presence of a face (RR Faces) predicted higher language comprehension scores, explaining over 7% of the variance. Higher Centre of Recurrence Mass for chair slides (CORM Chairs), which indicated greater proportion of revisits to selected objects later in the trial in the absence of a face, was associated with both higher comprehension and production. The latter predicted a relatively large proportion of the variance (approximately 16%–21%) in follow-up language measures. The two RQA measures, combined with family income, predicted approximately 40% of the variance in receptive and 30% in expressive language despite the sample being predominantly bilingual and composed of users of multiple languages.

Previous work on visual attention and language development often focused on the idea that greater attention to faces should be beneficial for infants' development, as it provides them with more opportunities for face-to-face social interactions and establishing communicative situations. While we do not dispute the importance of sufficient attention to faces for learning language, we note that existing studies show a relatively low proportion of time spent looking at faces early in infancy (Deak et al., 2014), even during naturalistic social interactions (Niedźwiecka et al., 2017). Although looking at faces gradually increases throughout the first year of life (e.g. Frank et al., 2014), in more complex displays there is an age-related increase in attention to hands and decrease in looking at faces (Frank et al., 2011), observed also during naturalistic interactions (Yu & Smith, 2013, 2016). Altogether, when viewing complex displays in our task 6-month-olds spontaneously focused both on objects and faces, so disproportionately high attention to faces could be considered suboptimal for scanning the scene. Moreover, prolonged fixating on faces at a cost of reduced attention to other stimuli may reflect difficulties with attention disengagement, rather than strong preference for social stimuli. This idea is supported by longer looking at faces and reduced disengagement from them in studies of infant siblings of children with autism (Elsabbagh et al., 2013; Jones et al., 2018), which predicted lower effortful control at 3 years (Hendry et al., 2018). Crucially, at 6 months of age, lower attention to the eyes was associated with better expressive language in infants siblings of children with autism (Wagner et al., 2018). Our results are consistent with these reports, but as we did not directly measure disengagement alongside visual scanning, this idea requires further testing.

We used a relatively simple scanning task, where infants viewed an array of objects either containing a face or a control chair image. The same categories of objects were repeated upon subsequent trials. Previous work has demonstrated that the presence of a face considerably affects the rate of fixation recurrence, as infants not only spend the majority of time fixating the face image (see Figure 1), but also revisit that area throughout the trial, as captured by the RR on face slides. Additional analyses show that it was indeed the face stimuli that drove repeated and recurrent fixations (Supplementary Tables S1 and S3) and longer average and total fixations (Supplementary Tables S2 and S4) in the face displays. Moreover, infants visited significantly fewer objects in face visual scenes than in the chair ones (Supplementary Table S6). However, infants with lower RR in face slides, which was predictive of their better language scores, visited more objects, suggesting that a more developed ability to disengage from faces is beneficial for language development (Supplementary Table S12). In the absence of a face, however, infants visited a higher number of objects (see Supplementary Table S6). Nonetheless, infants who more selectively attended to a few of them (see Supplementary Section 14a), producing higher CORM values on chair slides (see example in Figure 2), achieved better language scores at 24 months. This highlights the importance of considering exploration strategies as an independent predictor from social attention. Moreover, their individual scanning patterns were more dissimilar when no face stimulus was present (see López Pérez et al., 2018). These differences in individual participants' scanning behaviour between task conditions are an important feature of the task itself. The use of a face-absent condition led to a greater variability of visual scene exploration in the absence of a strong exogenous attention cue (e.g. a face). This is an important point given that several studies to date focused on attention to faces as the key predictor of early language development, while the current study highlights the role of exploration strategies as an independent predictor from social attention.

It is important to clarify the choice of the CORM for chairs and the RR for faces over other parameters such as the CORM for faces and the RR for chairs. Initial studies on this data showed that the exogenous effect from faces led to high correlations between the RR and CORM (see example in Figure 1) and therefore CORM is not able to pick up any significant differences in the temporal patterns of exploration (for more details, see López Pérez et al., 2018). In the chair slides, these correlations although smaller are still high and significant. Putting both in the same model would result in a collinearity between predictors and we would need to reduce them to single one. The complementary model, using CORM for faces and RR for chairs, still predicted receptive and productive scores in PLS-4, but it explained less variance than the original model (see Supplementary Table S10). Therefore, we chose those parameters that have a theoretical basis. First, since attention to faces supports language development, in this type of slides we chose the RR because it is a much more informative descriptor for whether the attention is directed to faces or not. Second, in the absence of the face, it is possible to study the efficiency of visual exploration over the image. In this particular case, the RR would be less informative since it would quantify which object is more interesting for each infant. This information is less important since there is no supporting evidence that any of the present categories is relevant for language development. However, in this case, the CORM will depict differences in the temporal patterns of explorations and describe, for instance, if infants scan rapidly the scene followed by the selection and detailed scanning of objects of interest (López Pérez et al., 2018; Manyakov et al., 2018). Therefore, using RR for faces and CORM for chairs, we combine dynamical information on attention to faces and visual exploration strategies, respectively, both likely predictors of language outcomes. What is particularly important is that dynamic measures of early attention predict later language development independently of family socio-economic status and bilingual status. This is particularly important for research with highly diverse samples, where predictor measures often co-vary with SES or show bilingual advantage (Singh et al., 2014; Tsang et al., 2018). Thus, our data demonstrate the potential utility of dynamic measures of visual scanning for predicting language outcomes for multicultural samples and multisite studies.

Our study tested a highly unique and diverse sample, both in terms of socioeconomic background and ethnicity. Moreover, nearly 75% of participating families were bi- or multilingual, using more than 40 languages; thus, the sample posed several challenges in terms of testing, as well as creating a need for an eye-tracking paradigm that reduces ethnic and linguistic biases. While all these factors can be considered further challenges to the use of standard experimental paradigms and test batteries, they also provide a critical test of the utility of these methods in real-world settings. Thus, our study is a step in the direction of applying existing basic research to develop early screening methods suitable for multicultural and multiethnic samples, for example, in global health research.

Another unique feature of our study was the out-of-laboratory data collection, using mobile eye-tracking equipment. It was carried out in early intervention centres, which were frequented by a large proportion of families at risk of multiple deprivation. The eye-tracking study at T1 was organised as an attractive activity for parents and included a short generic presentation of infant eye-tracking behaviour, which many especially low-SES parents found very helpful and important for their understanding of their child. Our previous analyses showed that mobile, out-of-lab data collection can provide reliable eye-tracking data of comparable quality (Ballieux et al., 2016). The current study demonstrates further that such data, combined with robust analyses of visual behaviour, can be a powerful tool for both conducting basic research in difficult to recruit groups, as well as a means for testing novel screening tools for early developmental difficulties. This conclusion is strengthened by the fact that these results were obtained for a bilingual and multiethnic sample, which uses multiple languages at home. Thus, our study also highlights a potential avenue for conducting research on early predictors of language even in highly varied samples.

The results should, however, be interpreted with some caution, as we note some limitations. First, the data have been collected for a sample with high variability on many dimensions, thus we consider the effects robust. However, the relatively low sample size and low proportion of monolingual children suggest a need for replication. In addition, the strict thresholds used during data reduction led to high participant attrition. However, we were aiming to get as close as possible to the infants' real visual exploration strategies and therefore we chose those infants that provided sufficient amounts of data. Second, the RQA analysis relies on accepting certain parameters such as the radius size. A recent study has observed similar results when using a radius such as 64 or 80 pixels, but several relationships disappeared when a radius of 48 pixels was used, suggesting that value might be too small (Manyakov et al., 2018). We decided on using the same size as used in previous studies, which in our case was almost equal to the size of each object within our visual scenes (Anderson et al., 2013; Wu et al., 2014). However, a more systematic analysis of the effect of the radius size is needed to quantify how the RQA parameters vary depending on it, but also how it is related to different types of stimuli (e.g. pop-out vs dynamic videos). Finally, infants in this study were presented with static scenes (including static faces), rather than with dynamic displays, which better correspond to real life viewing conditions. We opted for this choice because at 6 months of age infant looking is predominantly driven by exogenous mechanisms (moving objects) in dynamic displays, thus it is less possible to reliably measure their endogenously driven looking (Wass & Smith, 2014). However, this leaves open the question of what scanning strategies are optimal in more dynamic environments such as social interactions involving gaze and language cues from parents as well actions on objects. Studies measuring distractibility in dynamic displays suggest that endogenous mechanisms play an important part in shaping infant attention skills and predict later language outcomes (Salley et al., 2013). Also, head-mounted eye-tracking studies have shown that 1-year-olds rarely look to the parent's face and eyes during interactions involving manual actions on objects, but rather infants and parents coordinate looking behaviour without gaze following by attending to objects held by themselves or the social partner (Yu & Smith, 2017). Studies using head cameras in home environments complement this view and show that during the first 2 years of life, infants' attention shifts from predominantly faces during the first year to predominantly hands during the second year of life (Fausey et al., 2016; Jayaraman et al., 2015). Our results partially support this strand of work by showing that less attention to faces in the second half of the first year of life in the presence of other objects might be beneficial for language development. However, although moments of looking at the parent's face might be rare, they enable episodes of mutual gaze between parent and infant which create opportunities to practice infants' attention disengagement skills (Niedźwiecka et al., 2017). Thus, further research is needed to apply dynamical measures of looking in combination with mobile eye-trackers to investigate how more dynamic and naturalistic environments influence visual exploration and its relation to later language outcomes.

Conclusion

We demonstrated that dynamic measures of visual scanning using Recurrence Quantification Analysis provide a powerful tool for quantifying infant attention to both social and non-social stimuli. We also show that measures of the efficiency of visual scanning likely predict language development independently of attention to faces. This may suggest a potential mechanism linking early selectivity in attention to objects of interest with receptive and productive language development on the eve of preschool age. Finally, our results reinforce the utility of mobile eye-tracking in combination with well-defined experimental tasks for developing potential early screening tools for infants and children at risk of developmental difficulties.

Acknowledgements

Data collection was supported by the Nuffield Foundation (PI: D.M.). The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy, and also works to build capacity in education, science and social science research. More information is available at www.nuffieldfoundation.org. We thank all participating families for their vital contribution, as well as management and staff in Children's Centres in Tower Hamlets and Newham (London, UK). We especially thank Sally Parkinson, Head of Commissioning in Newham, for helping with setting up partnerships with CCs, and Monica Forty and her team for the ongoing support in Children's services in Tower Hamlets. We also thank Ewa Haman and Karolina Mieszkowska for helpful comments on the initial analysis. We sadly state that Annette Karmiloff-Smith passed away before this study was completed. We acknowledge the influence of her thinking and theoretical views on the design and writing up of this study.

Author contributions

Corresponding authors:
David López Pérez: Formal analysis, Software, Visualization, Writing – original draft, Writing – review and editing
Przemyslaw Tomalski: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft, Writing – review and editing
Alicja Radkowska: Formal analysis, Writing – original draft, Writing – review and editing
Co-authors:
Haiko Ballieux: Investigation, Methodology, Software, Validation, Writing – review and editing
Derek G. Moore: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review and editing

Data availability statement

The Matlab scripts and fixation data used in this manuscript are openly available in github at https://github.com/Mirandeitor/Language_Paper.git. Language follow-up scores are not available at the moment since they are part of an ongoing analysis.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. Eye-tracking data analyses presented in this work were funded from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 642996 (BRAINVIEW). Part of the analysis was developed thanks to the Polish National Science Centre Grant 2012/07/B/HS6/01464 to P.T. and the funds of the Institute of Psychology, Polish Academy of Sciences.

Supplemental material

Supplemental material for this article is available online.

ORCID iDs

David López Pérez https://orcid.org/0000-0002-1235-6376
Przemysław Tomalski https://orcid.org/0000-0002-0390-5759
Haiko Ballieux https://orcid.org/0000-0002-0933-1864

References

Anderson, N. C., Bischof, W., Laidlaw, K., Risko, E., & Kingstone, A. (2013). Recurrence quantification analysis of eye movements. *Behavior Research Methods*, *45*(3), 842–856.

Ballieux, H., Tomalski, P., Kushnerneko, E., Johnson, M. H., Karmiloff-Smith, A., & Moore, D. G. (2016). Feasibility of undertaking off-site infant eye-tracking assessments of neuro-cognitive functioning in early-intervention centres. *Infant and Child Development*, *25*(1), 95–113. https://doi.org/10.1002/icd.1914

Beier, J. S., & Spelke, E. S. (2012). Infants' developing understanding of social gaze. *Child Development*, *83*(2), 486–496. https://doi.org/10.1111/j.1467-8624.2011.01702.x

Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Development of audition and vision during the first year of postnatal life: A methodological overview* (pp. 253–300). Ablex.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*, 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., & Blaga, O. (2004). The developmental course of habituation and preschool outcome. *Infancy*, *5*, 1–38.

Comishen, K. J., Bialystok, E., & Adler, S. A. (2019). The impact of bilingual environments on selective attention in infancy. *Developmental Science*, *22*(4), e12797. https://doi.org/10.1111/desc.12797

Deak, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, *17*, 270–281.

Department for Communities and Local Government. (2010). The English Index of Multiple Deprivation (IMD) 2015 – Guidance. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/6222/1871538.pdf

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

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010). *Head-mounted eye-tracking of infants' natural interactions*. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications – ETRA '10. https://doi.org/10.1145/1743666.1743671

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170. http://doi.org/10.1016/j.cognition.2008.11.010

Frank, M. C., Vul, E., & Saxe, R. (2011). Measuring the development of social attention using free-viewing. *Infancy*, *17*(4), 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology*, *118*, 13–26. http://doi.org/10.1016/j.jecp.2013.08.012

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*(5), 550–562. https://doi.org/10.1080/15250000903144199

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science*, *13*(6), 839–848. https://doi.org/10.1111/j.1467-7687.2009.00945.x

Hains, S. M., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development*, *67*(5), 1940. https://doi.org/10.2307/1131602

Hendry, A., Jones, E. J., Bedford, R., Gliga, T., Charman, T., & Johnson, M. H. (2018). Developmental change in look durations predicts later effortful control in toddlers at familial risk for ASD. *Journal of Neurodevelopmental Disorders*, *10*(1). https://doi.org/10.1186/s11689-017-9219-4

Hessels, R. S., Andersson, R., Hooge, I. T., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, *20*(6), 601–633. https://doi.org/10.1111/infa.12093

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. (2016). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*, 1802–1823. https://doi.org/10.3758/s13428-016-08201

Hunnius, S., Geuze, R. H., & van Geert, P. (2006). Associations between the developmental trajectories of visual scanning and disengagement of attention in infants. *Infant Behavior & Development*, *29*(1), 108–25. https://doi.org/10.1016/j.infbeh.2005.08.007

Irwin, D. E. (1992). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146–165). Springer.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLOS ONE*, *10*(5), Article e0123780.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1–2), 1–19. https://doi.org/10.1016/0010-0277(91)90045

Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, *50*, 169–179. https://doi.org/10.1016/j.neubiorev.2014.10.009

Jones, E., Dawson, G., & Webb, S. (2018). Sensory hypersensitivity predicts enhanced attention capture by faces in the early development of ASD. *Developmental Cognitive Neuroscience*, *29*, 11–20. https://doi.org/10.1016/j.dcn.2017.04.001

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

López Pérez, D., Radkowska, A., Rączaszek-Leonardi, J., Tomalski, P., & Team, T. T. (2018). Beyond fixation durations: Recurrence quantification analysis reveals spatiotemporal dynamics of infant visual scanning. *Journal of Vision*, *18*(13), 5. https://doi.org/10.1167/18.13.5

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, *128*, 85–93.

Manyakov, N. V., Bangerter, A., Chatterjee, M., Mason, L., Ness, S., Lewin, D., & Pandina, G. (2018). Visual exploration in autism spectrum disorder: Exploring age differences and dynamic features using recurrence quantification analysis. *Autism Research*, *11*(11), 1554–1566. https://doi.org/10.1002/aur.2021

Marino, C., & Gervain, J. (2019). The novelty effect as a predictor of language outcome. *Frontiers in Psychology*, *10*, 258. https://doi.org/10.3389/fpsyg.2019.00258

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, *21*, 373–377.

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*(4), 643–655. https://doi.org/10.1037/0012-1649.42.4.643

Niedźwiecka, A., Ramotowska, S., & Tomalski, P. (2017). Mutual gaze during early mother-infant interactions promotes attention control development. *Child Development*, *89*(6), 2230–2244. https://doi.org/10.1111/cdev.12830

Parise, E., Handl, A., Palumbo, L., & Friederici, A. D. (2011). Influence of eye gaze on spoken word processing: An ERP study with infants. *Child Development*, *82*(3), 842–853. https://doi.org/10.1111/j.1467-8624.2010.01573.x

Parise, E., Reid, V. M., Stets, M., & Striano, T. (2008). Direct eye contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, *3*(2), 141–150. https://doi.org/10.1080/17470910701865458

Parsons, J. P., Bedford, R., Jones, E. J. H., Charman, T., Johnson, M. H., & Gliga, T. (2019). Gaze following and attention to objects in infants at familial risk for ASD. *Frontiers in Psychology*, *10*, 1799. https://doi.org/10.3389/fpsyg.2019.01799

Perone, S., & Spencer, J. P. (2013). Autonomous visual exploration creates developmental change in familiarity and novelty seeking behaviors. *Frontiers in Psychology*, *4*, 648. https://doi.org/10.3389/fpsyg.2013.00648

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

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

Salley, B., Panneton, R. K., & Colombo, J. (2013). Separable attentional predictors of language outcome. *Infancy: The Official Journal of the International Society on Infant Studies*, *18*(4), 462–489. doi.org/10.1111/j.1532-7078.2012.00138.x

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, *108*, 303–319. https://doi.org/10.1016/j.cognition.2008.02.009

Singh, L., Fu, C. S., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., . . . Rifkin-Graboi, A. (2014). Back to basics: A bilingual advantage in infant visual habituation. *Child Development*, *86*(1), 294–302. https://doi.org/10.1111/cdev.12271

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Pergamon Press.

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2014). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/s0305000914000725

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

Wagner, J. B., Luyster, R. J., Moustapha, H., Tager-Flusberg, H., & Nelson, C. A. (2016). Differential attention to faces in infant siblings of children with autism spectrum disorder and associations with later social and language ability. *International Journal of Behavioral Development*, *42*(1), 83–92. https://doi.org/10.1177/0165025416673475

Wass, S. V., & Smith, T. J. (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. *Infancy*, *19*(4), 352–384. https://doi.org/10.1111/infa.12049

Wu, D. W. L., Anderson, N. C., Bischof, W., & Kingstone, A. (2014). Temporal dynamics of eye movements are related to differences in scene complexity and clutter. *Journal of Vision*, *14*(9), 8.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 229–248. https://doi.org/10.1080/15250000802004437

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development*, *88*(6), 2060–2078. https://doi.org/10.1111/cdev.12730

Zimmerman, I., Steiner, V., & Pond, R. (2002). *Preschool language scale* (4th ed.). The Psychological Corporation.