

Beyond fixation durations: Recurrence quantification analysis reveals spatiotemporal dynamics of infant visual scanning

David López Pérez

Neurocognitive Development Lab, Faculty of Psychology,
University of Warsaw, Poland



Alicja Radkowska

Neurocognitive Development Lab, Faculty of Psychology,
University of Warsaw, Poland

Joanna Rączaszek-Leonardi

Faculty of Psychology, University of Warsaw, Poland

Przemysław Tomalski

Neurocognitive Development Lab, Faculty of Psychology,
University of Warsaw, Poland



The TALBY Study Team

TALBY Study Team: Haiko Ballieux, Elena Kushnerenko,
Mark. H. Johnson, Annette Karmiloff-Smith, Deirdre
Birtles & Derek G. Moore

Standard looking-duration measures in eye-tracking data provide only general quantitative indices, while details of the spatiotemporal structuring of fixation sequences are lost. To overcome this, various tools have been developed to measure the dynamics of fixations. However, these analyses are only useful when stimuli have high perceptual similarity and they require the previous definition of areas of interest (AOIs). Although these methods have been widely applied in adult studies, relatively little is known about the temporal structuring of infant gaze-foraging behaviors such as variability of scanning over time or individual scanning patterns. Thus, to shed more light on the spatiotemporal characteristics of infant fixation sequences we apply for the first time a new methodology for nonlinear time-series analysis—the recurrence quantification analysis (RQA). We present how the dynamics of infant scanning varies depending on the scene content during a “pop-out” search task. Moreover, we show how the normalization of RQA measures with average fixation durations provides a more detailed account of the dynamics of fixation sequences. Finally, we link the RQA measures of temporal dynamics of scanning with the spatial information about the stimuli using heat maps of recurrences without the need for defining a priori AOIs and present how infants’ foraging strategies are driven by the image content. We conclude from our findings that the RQA methodology has potential applications in the analysis of the temporal dynamics of infant

visual foraging offering advantages over existing methods.

Introduction

As naive learners, during the first year of life, infants face a number of challenges in building their knowledge of the structure of the surrounding environment (Aslin, 2013). From birth, they rely on visual information processing and limited attention biases to guide their learning, most notably a bias towards face-like patterns (Johnson, Dziurawiec, Ellis, & Morton, 1991; for a review see Johnson, Senju, & Tomalski, 2015). Initially, reflexive saccades (i.e., eye movements occurring in response to the sudden onset of a peripheral stimulus) dominate in the first two postnatal months (Richards, 2008) and young infants have difficulties with disengaging their gaze from any stimulus (e.g., Johnson, Posner, & Rothbart, 1991). At around two months of age, infants begin to show smooth visual tracking, although their eye movements still lag behind the movement of the stimulus (Johnson & de Haan, 2015, p. 93). Moreover, the cortical areas that control voluntary saccade movements in the brain continue to develop until the sixth postnatal month, which allows infants to control attention-directed voluntary saccades. As visual scanning of the environment becomes

Citation: López Pérez, D., Radkowska, A., Rączaszek-Leonardi, J., Tomalski, P., & the TALBY Study Team. (2018). Beyond fixation durations: Recurrence quantification analysis reveals spatiotemporal dynamics of infant visual scanning. *Journal of Vision*, 18(13):5, 1–17, <https://doi.org/10.1167/18.13.5>.

<https://doi.org/10.1167/18.13.5>

Received March 23, 2018; published December 7, 2018

ISSN 1534-7362 Copyright 2018 The Authors



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

more elaborate throughout the first year of life, infant looking is increasingly driven by their prior knowledge and less by low-level salience (Frank, Vul, & Johnson, 2009). Finally, the brain pathways that control smooth pursuit movements continue to develop until the second year of life (Richards, 2008).

As the research on infant visual cognition investigates component processes in increasing detail, the field has moved away from the macrostructure of overall looking times and is focused on the microstructure of individual fixations measured with eye-tracking (Aslin, 2012). Looking-time measures were introduced by Fantz (1963, 1964), who demonstrated that newborns have an organized and selective looking behavior despite poor visual acuity. However, these measures do not allow exploring the microdynamics of visual and cognitive processing. Therefore, a growing body of literature is focusing on the analysis of fixation durations (FDs) as indicators of cognitive processing, which represents how long a fixation remains in a particular location (Saez de Urabain, Nuthmann, Johnson, & Smith, 2017). FDs are typically combined in order to compare their average and the total time spent looking between images or particular areas of interest (AOIs) of an image (e.g., Tenenbaum, Shah, Sobel, Malle, & Morgan, 2012; Tomalski et al., 2013). Hence, FDs allow the calculation of global measures of cumulative and average looking times, which have been successfully related to a range of aspects of cognitive development (Aslin, 2012), individual differences in temperament and attention (Papageorgiou et al., 2014), or efficiency of information processing (Colombo & Mitchell, 2009). However, standard looking-time analysis typically produces only a global quantitative description of the eye-tracking data and does not provide information about the temporal structuring of gaze-foraging behaviors such as within-subject variability of scanning paths over time or individual scanning patterns.

Recently, in the adult eye-tracking research community there has been an increased interest in the measurement of temporal dynamics of fixations to analyze the *scanpaths* (i.e., trajectories of the eyes when scanning the visual field and analyzing any kind of visual information). In scanpath methods, images are divided into AOIs, which can either be designated feature regions within an image (e.g., body parts or objects) or can be created by simply binning (i.e., the process of combining a cluster of pixels into one single pixel) the image into a discrete number of regular bins (for a review, see Anderson, Anderson, Kingstone, & Bischof, 2015). A letter is then assigned to each region, and every eye fixation within that region is tagged with this identifier. Methods such as string-edit distance (e.g., Foulsham & Kingstone, 2013), linear distance algorithm (Henderson, Brockmole, Castelhamo, &

Mack, 2007), ScanMatch (Cristino, Mathôt, Theeuwes, & Gilchrist, 2010), or the MultiMatch analysis (Dewhurst et al., 2012) have proved to be useful for comparing scanpaths among participants (Anderson et al., 2015). Scanpath analyses have shown, for instance, that infants collect facial information more efficiently from upright faces than from inverted ones, an ability that gradually develops with age (Kato & Konishi, 2013), and that they scan faces of various ethnicities differently (e.g., Xiao, Xiao, Quinn, Anzures, & Lee, 2012). However, these analyses are generally constrained to comparisons for closely related stimuli or very similar layouts where particular points are easily comparable, making it difficult to generalize the scanpath measures over participants and stimuli (for a further discussion, see Anderson et al., 2015).

Scanpath and the majority of FD analyses require a prior definition of relevant AOIs, which is an important limitation. On one hand, AOIs are needed to link FDs with the spatial location of visual stimuli, while on the other hand, they are needed to create the different scanpaths. This also limits the analysis to independent and spatially separated areas where nearby fixations may potentially be wrongly classified into different AOIs (Anderson et al., 2015), or spatial information can be lost from areas that in some cases are not considered relevant to the analysis (e.g., background). Moreover, AOIs often differ in size and shape even if the stimuli are similar (Hessels, Kemner, van den Boomen, & Hooge, 2016a). This can be problematic for several reasons. First, it complicates the comparison between studies using similar stimuli (Hessels et al., 2016a). Second, it complicates the analysis of the temporal dynamics of fixations when the image is not considered as a whole but divided into independent areas. Finally, it requires the researcher to process the context of the image in order to select the AOIs.

A recent study has introduced a machine-learning approach to analyze eye-tracking data in a decontextualized manner (i.e., without the definition of any AOIs) and to overcome these limitations that the definition of AOIs introduces (Vallee, 2015). In this way, the eye-tracking data comprises the entire interaction of the user with the image. This is an important advantage over AOI-based studies because it does not suffer from the restrictions that AOIs introduce.

In this paper, we demonstrate that the investigation of scan patterns in infant eye-tracking data can be considerably enhanced by advanced methods developed for the study of dynamical properties of cognitive systems. Specifically, we measure the spatiotemporal properties of infant visual scanning during a face “pop-out” task using a technique known as recurrence quantification analysis (RQA; Zbilut, Giuliani, & Webber, 1998). RQA is an increasingly popular method

of analyzing dynamical changes of behavior in complex systems. In psychology, it has become an attractive tool revealing the complexity of human behavior across multiple timescales. It has been applied to measure the coupling of two behavioral signals from different interacting actors (e.g., Shockley, Santana, & Fowler, 2003; Richardson & Dale, 2005; Shockley & Turvey, 2005), behavioral coupling in parent–child interactions (López Pérez et al., 2017), including gaze coupling (Nomikou, Leonardi, Rohlfing, & Raczaszek-Leonardi, 2016) or stereotypical motor behavior of individuals with autism spectrum disorder (Romero, Fitzpatrick, Schmidt, & Richardson, 2016; Großekathöfer et al., 2017).

Several attempts have been made to apply RQA to study the dynamics of eye-tracking data (e.g., Richardson & Dale, 2005; Cherubini, Nüssli & Dillenbourg, 2010). Early attempts focused on analyzing visual scanning as a sequence of fixations on a few AOIs and used this data in subsequent RQA analysis to find common dynamics. However, only recently has RQA been extended to characterize the temporal dynamics of fixation coordinates (Anderson, Bischof, Laidlaw, Risko, & Kingstone, 2013; Wu, Anderson, Bischof, & Kingstone, 2014; Demiralp, Cirimele, Heer, & Card, 2017). Anderson et al. (2013) used RQA to study the individual differences in scan patterns under natural (i.e., unrestricted scene-viewing task) versus gaze-contingent (i.e., the fixation field changes depending on the participant's eye movements) viewing conditions. Wu et al. (2014) expanded the previous study and explored the relationship between RQA parameters versus the complexity and clutter in visual scenes. Both studies showed that local and global temporal properties of fixations can be described by a handful of parameters and that these parameters were sensitive to the type of scene or the scene viewing conditions. Additionally, it was shown that these measures had a clear interpretation within the context from which they were extracted. For instance, a higher percentage of fixations that form repeated trajectories of fixations (determinism; DET) and a higher percentage of fixations that form consecutive fixations in the same image areas (laminarity; LAM) were found when certain regions of the scene were explored in more detail. Additionally, the percentage of fixations that are part of areas previously fixated (recurrence rate; RR) varied in relation to the image content of the scene, and the global patterns of refixations (center of recurrence mass; CORM) changed depending on the viewing conditions (Anderson et al., 2013).

The advantage of RQA in comparison to standard looking-time analysis is the possibility of disentangling local from global gaze behavior in one go (Anderson et al., 2013). Variations in the DET and the LAM reflect changes in the local gaze behavior. Moreover, changes

in the RR and the CORM are related to changes in the global gaze behavior. RQA also allows to analyze fixation data in a decontextualized manner without the definition of any areas of interest. This is achieved first by extracting the RQA parameters and, second by back-projecting these measures onto the image space (Anderson et al., 2013). More importantly, RQA allows the analysis of the temporal dynamics of a single scanpath while one of the major disadvantages of scanpath analyses is that, until recently, two scanpaths were needed to perform any comparison (Anderson et al., 2015). Altogether, this allows for differences in the temporal dynamics to be captured, to provide in-depth depiction of what drives infants' attention, and to explore how these processes evolve over time, which is something that cannot be investigated with traditional high-level measures.

To our knowledge, RQA has not been used to analyze the temporal dynamics of infant eye-tracking data, which is often complicated. Infants do not sit still, their oculomotor system undergoes rapid development and task performance can considerably vary from one subject to another, and their attention span is shorter than adults, which leads to lower data quality and reduced number of valid fixations (e.g., Holmqvist, Nyström, & Mulvey, 2012; Wass, Smith, & Johnson, 2012).

The purpose of our study was to apply for the first time the RQA methodology to analyze infant eye tracking (ET) data during a face pop-out task in a large sample of 6- to 7-month-old infants. In adult studies, the content of a visual scene (e.g., images of landscapes vs. interiors) is related to both global (RR and CORM) and local temporal gaze patterns (LAM and DET) (Anderson et al., 2013). First, we assessed the usefulness of the RQA measures in infant ET data and tested whether the presence of a human face in a visual scene with objects affected the global and the local temporal gaze patterns. Second, we normalized the RQA with average FDs for each participant to control for individual differences in FDs, which may reflect variability in infant vigilance or processing speed. Third, we independently tested without the use of predetermined AOIs, which areas in the visual image are related to the face versus chair differences in RQA measures by back-projecting the RQA data onto the image. Finally, we computed the similarity between the infants' individual RQA maps and showed how infants' foraging strategies are driven by the image content.

Overall, we expected that participants would fixate on faces more often than other objects (Gliga, Elsabbagh, Andravizou, & Johnson, 2009), showing higher RR in visual scenes with a face than scenes without them (i.e., chair scenes). Additionally, because faces attract and hold attention longer in comparison to other objects (e.g., Langton, Law, Burton, &

Schweinberger, 2008), participants would show higher LAM in face scenes than in chair scenes. Since faces attract infants' attention, there is a higher chance to observe repeated scanning patterns from the face to different objects and back to the face. Therefore, we expected higher DET in the face scenes than the chair ones, which were higher object exploration and thus, lower DET is expected. Finally, we anticipated longer temporal gaps between refixations in the visual scenes with faces (i.e., higher CORM) due to the exogenous cueing of faces, which would attract infants' attention more often during the trial.

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 multiple deprivation nationwide. 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 and eighty-three infants ($N = 183$) were recruited to the study and 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 since they did not produce enough ET data (for inclusion criteria see Eyetracking data preprocessing section below). The final sample consisted of 109 infants aged between 6 months 1 day and 7 months 30 days ($M = 207.91$ days; $SD = 21.59$), with 40 girls (36.7%) and 69 boys (63.3%). Reflecting the mixed ethnic composition of East London, the final sample comprised 25 (25.9%) Caucasian, 14 (11.5%) Afro-Caribbean, 45 (34.5%) Asian-Indian, and 25 (28.2%) mixed ethnicity infants, originating from different countries and language groups. All participants included in the sample were born full-term (36–42 weeks gestational age), without 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 sixteen

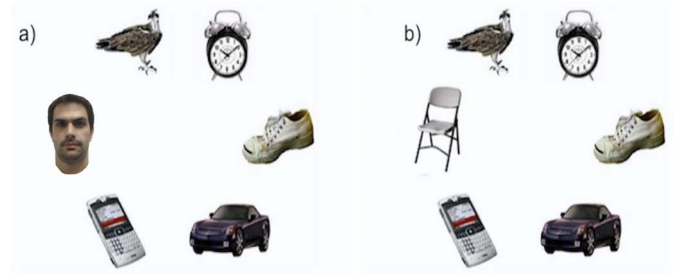


Figure 1. Example of the stimuli presented. Five objects, common in the 10 scenes, were varied examples from the categories of shoes, cars, mobiles, birds, and clocks. The remaining object was selected from two categories of objects: faces (a) or chairs (b). The human face is an example of what the original stimulus looks like. Written consent was given by the first author of this publication to use the image.

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 Tower Hamlets 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.

Face pop-out task

We used a modified (Ballieux et al., 2016) face pop-out task (Gliga et al., 2009), in which infants freely viewed visual scenes of six colored objects on a white background (Figure 1a and 1b). Ten visual scenes were created, with each 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 s. There were two different pseudorandom orders of presentation, in which the 10 scenes were presented in two blocks, with the block order counterbalanced between subjects.

Data acquisition

The data were acquired using a portable kit, which contained a 17 in. eye-tracker with integrated monitor

(T120; Tobii Technology, Inc., Reston, VA) and a portable Ergotron MX desk mount arm (Ergotron, St. Paul, MN) 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.

Eye-tracking data preprocessing

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. Additionally, participants that did not provide at least three valid trials for each visual scene type were rejected from the analysis sample. 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, Niehorster, Kemmer, & Hooge, 2016b). This algorithm can detect fixations in noisy data, which makes it suitable for infant research in which data quality is generally poorer than adult studies (Hessels, Andersson, Hooge, Nyström, & Kemner, 2015) and especially suitable for our dataset, 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., 2016b). For the Steffen interpolation, we used a window of 100 ms and an interpolation edge of two samples (i.e., 16.66 ms). In the k -means clustering, we applied a sample-by-sample analysis; a clustering window size of 200 ms; downsampling at 60, 30, and 15 Hz; and a clustering cutoff of 2 times the standard deviation above the k -means weights. Finally, all those fixations that had a minimum duration of 40 ms were considered valid, and we merged fixation candidates that were less than 0.7° apart and separated by less than 30 ms.

Recurrence quantification analysis of eye-tracking data

We used RQA to analyze the spatiotemporal characteristics of fixation sequences in pre-processed (fixation-filtered) ET data, an analysis developed to characterize the gaze patterns of single observers (Anderson et al., 2013; Wu et al., 2014). We estimated the RQA parameters using a Matlab Toolbox found in: <http://barlab.psych.ubc.ca/research/> (see detailed de-

scription in Anderson et al., 2013). Below, we outline the definitions of these parameters:

- **RR**: Represents the percentage of recurrent fixations (i.e., number of refixations in previously fixated image areas).
- **LAM**: Percentage of recurrent points that form vertical lines (i.e., sequential fixations that consecutively fixate on the same location). The minimum length of these structures that should be considered was set to two (i.e., two consecutive fixations).
- **DET**: The percentage of recurrent points that fall on diagonal lines in the RQA plot (i.e., specific sequences of fixations or scan paths that repeat). The minimum length of these structures that should be considered was also set to two.
- **CORM**: Measures the temporal distribution of recurrences by comparing the recurrences close to the diagonal line in the RQA plot with those further away from it. Small values indicate that recurrences occur at close proximity in time, while high values represent large temporal intervals between recurrences.

Figure 2 shows an example of an RQA plot (Figure 2a) for the fixation-filtered data, for which the original scanpath is plotted in Figure 2b, with individual fixations marked by circles. A red dot is drawn in the plot when two fixations fall within the distance radius (e.g., 64 pixels). The number of recurrences (i.e., each red dot) in the RQA plot increases when an area of the image is repeatedly revisited. For instance, if a participant explores the entire scene in detail then recurrences will be sparser in the plot (Figure 2a). These recurrences can be isolated as single points or occur closer together forming different structures. For instance, during the exploration of the scene, there can be a particular fixation pattern that repeats (black ellipses in Figure 2a). In this case, the particular pattern in Fixations 7, 8, and 9 are repeated in Fixations 16, 17, and 18, as well as in Fixations 27, 28, and 29. This type of diagonal structures contributes to the DET measure. Likewise, there could be an instant in which a specific, previously fixated area is consecutively fixated, forming vertical or horizontal lines in the plot (Fixations 18, 19, and 20 inside the blue square in Figure 2a). These refixations at the same location contribute to the LAM. In this paper, fixations were recurrent if they fell within a 64-pixel radius (Anderson et al., 2013; Wu et al., 2014) of another fixation, which corresponds to a 3.16° of visual angle. This value also corresponds to the approximate size of the objects within the stimuli.

Normalized RQA measures

To account for the variability in FDs in infants we normalized the RQA measures to consider individual

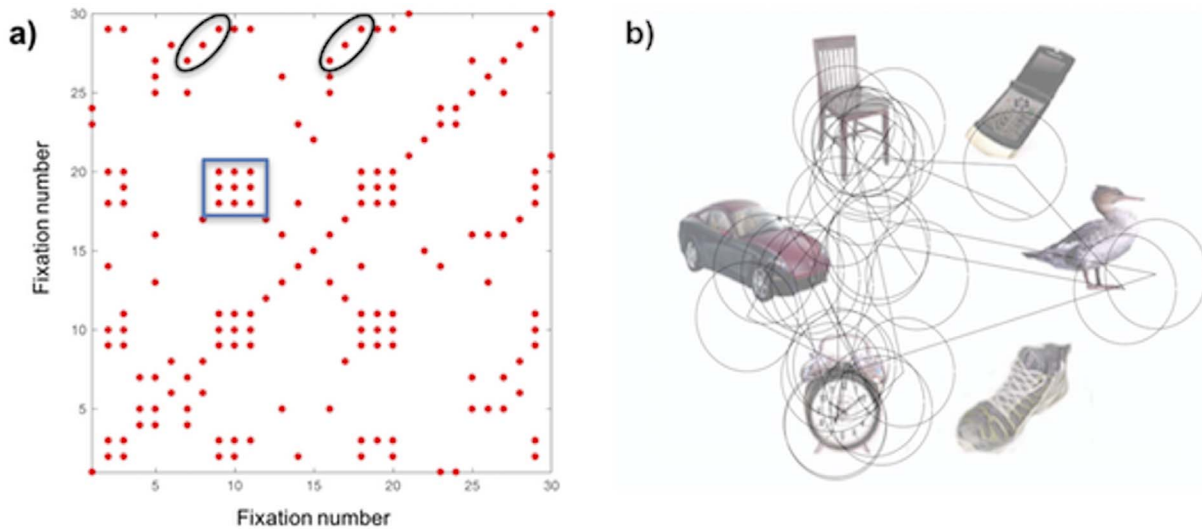


Figure 2. Example of an RQA plot (a) and its original fixation coordinates for a visual scene containing a chair (b). The black ellipse represents deterministic fixations sequences where Fixations 7, 8, and 9, recurred with Fixations 16, 17, and 18, as well as Fixations 27, 28, and 29. The blue square shows consecutive patterns of fixations (Fixations 18, 19, and 20) at the same location. This plot comes from the data of one participant.

differences. This normalization process can be done by redefining the calculation of recurrence to account for FDs (see the redefined RQA measures appendix in Anderson et al., 2013). We followed this procedure for all the RQA measures.

Statistical analysis of the RQA measures

Dependent *t* tests were computed to test differences between all the RQA measures of visual scenes with faces and chairs. Order effects were tested using a 2×3 ANOVA with image type (visual scenes with faces and chairs) and repetitions (acquisition order) as within factors. No interaction effects were found between image type and repetitions order in any of the RQA analysis (all *ps* > 0.05).

Associations between average and total FDs and RQA measures

We tested whether the RQA parameters are related to the information provided by classical measures using fixation duration data. First, we correlated these recurrence measures with total duration of fixations per AOI in both face and chair scenes. Thus, a single elliptical AOI was defined to cover each object in both types of scenes. All the fixations falling within this AOI were used to compute the total fixation durations for each object. Next, we separately correlated average fixation durations for face and chair scenes with the recurrence measures extracted from the fixation-filtered data.

RQA heat maps

Previously calculated RQA measures were projected back onto the image space in order to link the temporal dynamics with the spatial information in the image. These RQA heat maps display the complex dynamics of visual scanning measured with RQA in relation to the original image (Anderson et al., 2013). In this study, we produced three kinds of maps:

- **Recurrence heat map:** Represents those areas of the scene that were refixated by the infants. Thus, all recurrences (see red dots from the RQA plot in Figure 2a) were back-projected onto the image space (see Figure 3a and 3b for an example of the original scanpath and the corresponding recurrence heat map).
- **Determinism heat map:** Contains information about areas of the image that were part of repeated scan paths. It was obtained by back-projecting all those recurrence points that were part of diagonal lines in the RQA plot (see Figure 3c for an example determinism heat map).
- **Laminarity heat map:** Provides the locations in the image that are repeatedly fixated by back-projecting all those recurrences that were part of vertical lines (see Figure 3d for an example laminarity heat map).

The same radius (see the Recurrence quantification analysis of eye-tracking data section) that was used to calculate the RQA parameters was used to back-project the measures into image space. The final heat maps were filtered with a 2D Gaussian smoothing kernel and overlaid on the original image. If participants fixated

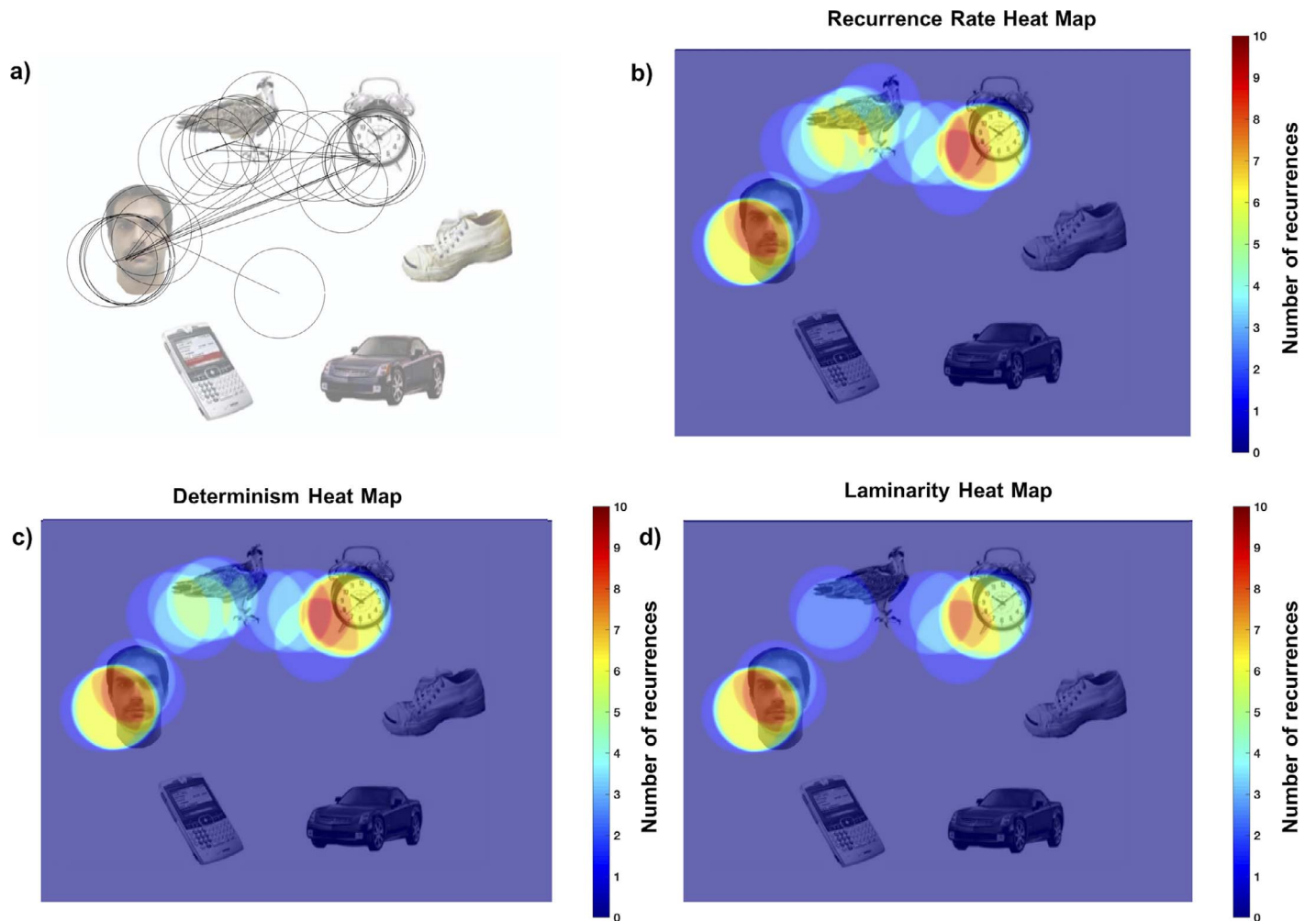


Figure 3. Examples of an individual recurrence rate heat map for one scene in one participant: (a) represents the original fixation sequence, (b) its equivalent recurrence heat map, (c) the determinism heat map, and (d) the laminarity heat map. Higher intensities (i.e., warmer colors) reflect areas that are more refixated. This plot comes from the data of one participant. The human face is an example of what the original stimulus looks like. Written consent was given by the first author of this publication to use the image.

the face more frequently than other objects, we expected to find a higher number of recurrences located on faces.

Although the heat maps allow us to interpret the RQA measures, we additionally extracted the total number of fixations that are part of the recurrence, laminarity and determinism measures for each object, within the five scenes of each type (face vs. chair scenes). Individual heat maps were added together to produce a total single heat map for each face and chair scene. A single mask was used to quantify the total number of recurrences located at each object for each recurrence measure used to draw the heat maps. Finally, the values for each type of scene were averaged. We expected a higher total number of recurrences for faces in comparison to the other objects in face scenes and more equally distributed recurrences in the chair ones.

Next, to determine the extent to which individuals look at the same location in the scenes we applied pairwise comparison between the heat maps (Le Meur & Baccino, 2012; Kennedy et al., 2017). Thus, we vectorized each heat map (i.e., converted to a single column of values) and compared them with Pearson's correlations. For each face or chair scene, we correlated each individual heat map with the heat maps of the rest of participants within the same scene (e.g., the heat map of one participant in scene Face 1 was correlated with all the remaining individuals; heat maps in scene Face 1). Mean similarity was then calculated for each scene. To account for the non-normal distribution of correlation coefficients, paired t tests were performed on the Fisher r -to- Z transformed correlation coefficients. Finally, to control for false positive similarity results that could arise from the shared structure of visual stimuli across trials we compared each face heat map to a random model where the infants' scanning

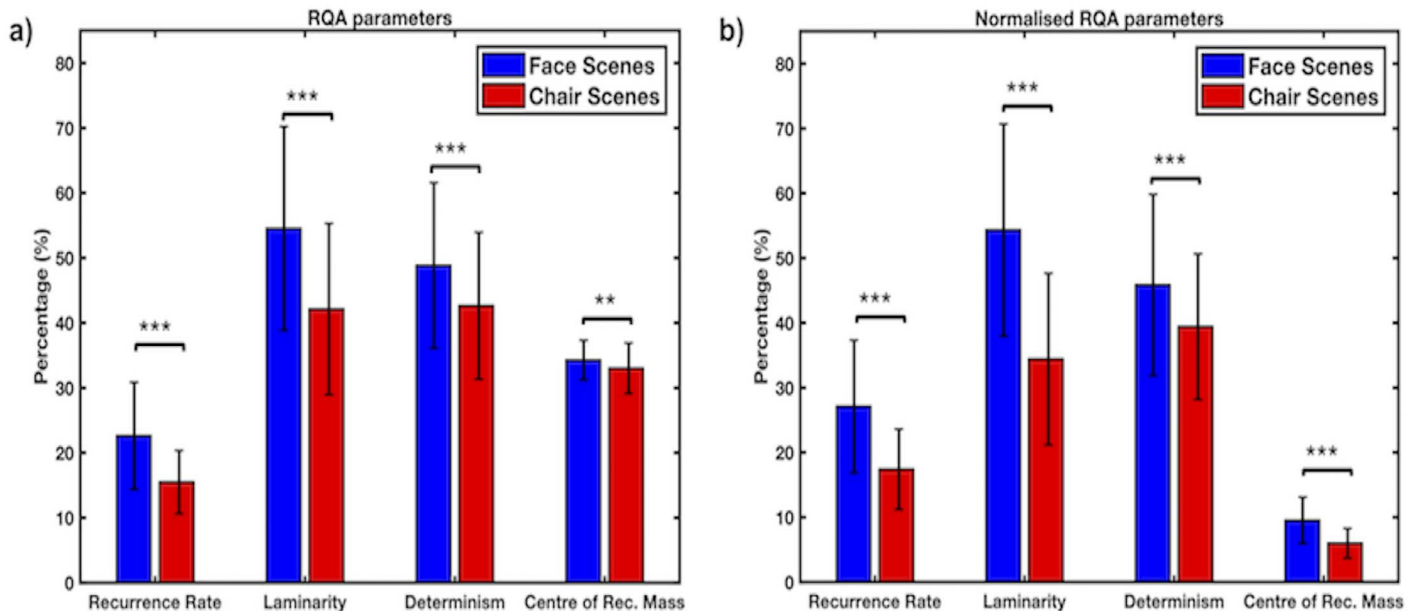


Figure 4. Average values for RQA parameters (a) and normalized RQA parameters (b) of infant eye-tracking data for visual scenes with faces (blue) or chairs (red; ** $p < 0.01$, *** $p < 0.001$). The error bars represent the standard deviation.

behavior was scrambled across the scenes. Faces have a strong exogenous effect that can drive fixation sequences to them, and therefore, we expected infants to show a higher similarity of fixations sequences in face versus chair scenes versus random model. We followed this procedure for the heat maps derived from each RQA measure but the CORM.

Results

Differences in RQA measures for visual scenes with face versus chair

RQA was used to study the individual spatiotemporal properties of fixations within face and chair visual scenes. At the group level, all RQA parameters were significantly higher for visual scenes with a face compared to those with a chair: RR, $t(108) = -8.95$, $p < 0.001$, $d = -1.05$; CORM, $t(108) = -3.05$, $p = 0.003$, $d = -0.20$; LAM, $t(108) = -8.27$, $p < 0.001$, $d = -0.85$; and DET, $t(108) = -4.18$, $p < 0.001$, $d = -0.51$ (see Figure 4a for group averages). This suggests that when a face is present, infants scan the scene and often return to parts previously fixated (higher RR). Areas in the scene that have been initially fixated are revisited more often later in the trial (higher CORM). Specific sequences of fixations or scan paths frequently repeat (higher DET) and there is more detailed scanning in specific regions of the scene that infants scan longer (higher LAM).

To correct for the large individual differences in average fixations durations (FDs) in infants, we normalized and computed again the RQA parameters. At the group level, the normalized measures were significantly different between chair and face visual scenes in RR, $t(108) = 10.16$, $p < 0.001$, $d = -1.14$; CORM, $t(108) = -9.89$, $p < 0.001$, $d = 0.51$; LAM, $t(108) = -14.24$, $p < 0.001$, $d = -1.33$; and DET, $t(108) = -4.45$, $p < 0.001$, $d = -0.50$ (see Figure 4b for group averages). The differences in mean values between parameters of both visual scenes increased with the normalization of the RQA measures (see summary in Table 1).

Associations between total fixation durations and the RQA parameters

We tested to what extent the RQA parameters are related to the information provided by classical measures using fixation duration data. First, we separately correlated individual total fixation durations per AOI with the RQA measures. We observed consistent significant correlations between the total fixation durations on the face AOI and three of the recurrence measures for face scenes: RR, $r(108) = 0.67$, $p < 0.001$; LAM, $r(108) = 0.63$, $p < 0.001$; DET, $r(108) = 0.43$, $p < 0.001$. This suggests that the spatiotemporal properties of fixation sequences might be driven by the presence of a face and hence, the strong relationship with cumulative fixation times in the face AOI. Additionally, the fixation duration on the clock AOI in the visual scenes with chairs showed low significant

RQA parameters

	RR (%)	RR _{NORM} (%)	LAM (%)	LAM _{NORM} (%)	DET (%)	DET _{NORM} (%)	CORM	CORM _{NORM}
Faces	22.61 (8.25)	27.08 (10.24)	54.50 (15.69)	54.31 (16.38)	48.80 (12.75)	45.83 (13.97)	34.26 (3.06)	9.51 (3.58)
Chairs	15.47 (4.86)	17.39 (6.21)	42.11 (13.18)	34.40 (13.23)	42.62 (11.30)	39.40 (11.25)	33.01 (3.88)	5.97 (2.28)

Table 1. This table summarizes the mean RQA parameters before and after the normalization (*SD* in parentheses, subscript NORM indicates the normalized parameter).

correlations with two of the recurrences measures: RR, $r(108) = 0.32$, $p = 0.001$; LAM, $r(108) = 0.24$, $p = 0.01$. No other significant correlations ($p > 0.05$) were found between the RQA measures and the rest of the objects (see Supplementary File S1 for detailed statistics in each of the remaining AOI).

Associations between average fixation durations and the RQA parameters

Next, we ran Pearson correlations between the average fixation durations within faces and chair slides with their corresponding RQA parameters (see Supplementary File S1 for FD group averages). In the face scenes, we observed significant correlations of average FDs for RR, $r(108) = 0.40$, $p < 0.001$; LAM, $r(108) = 0.27$, $p = 0.004$; and DET, $r(108) = 0.21$, $p = 0.02$, but not for the CORM, $r(108) = -0.11$, $p = 0.25$. Likewise, in the chair scenes average fixation durations correlated with RR, $r(108) = 0.28$, $p = 0.003$ and LAM, $r(108) = 0.24$, $p = 0.01$, but not with DET, $r(108) = 0.14$, $p = 0.14$ or CORM, $r(108) = -0.12$, $p = 0.19$.

RQA heat maps

Next, we generated the heat maps in order to link some of the recurrence measures (RR, DET, and LAM) with specific regions of the image. Figure 5 shows an example of the total heat map across participants for one particular scene with a face and its corresponding scene with a chair. It is observed that in the visual scene with a face, all the recurrences (Figure 5a), recurrences that contribute to the laminarity (Figure 5c), or to the determinism (Figure 5e), are mainly positioned over the face image, with fewer recurrences falling on the other objects. On the other hand, in the visual scene with a chair, all the recurrences (Figure 5b), laminar recurrences (Figure 5d), and deterministic recurrences (Figure 5f), are more evenly distributed among a higher number of objects indicating greater exploration of the scene.

Next, we extracted the total number of recurrent fixations for each object within each scene (see Supplementary File S1 for the descriptive statistics of recurrences in each type of heat maps). Tables 2 and 3

summarize the percentage of recurrent fixations for each type of heat map. As we observe, when a face is present, infants fixate predominantly on it (~52%), specific sequences of fixations or repeated scanpaths are influenced by faces (~58%), and the detailed scanning in specific regions focuses mainly on faces (~63%). On the other hand, when a face is absent, infants tend to explore other objects, which are visited approximately twice as often in chair scenes than the face scenes. Interestingly, in the visual scenes containing chairs, clocks were visited more often than the remaining objects.

Similarity in scanning patterns

Finally, we investigated to what extent the presence of a face influences scanning patterns, by comparing each infant's individual heat map with the heat maps of the other infants in the face and chair scenes. A significant main effect of scene type was found for the recurrence rate, $t(523) = 21.76$, $p < 0.001$, $d = 1.28$; laminarity, $t(466) = 21.04$, $p < 0.001$, $d = 1.38$; and determinism, $t(439) = 19.65$, $p < 0.001$, $d = 1.30$ heat maps (see Figure 6 for the average similarity for the chair and face scenes and the random model obtained from the heat maps of each RQA measure). The similarity in both types of scenes was significantly higher in comparison to the random model ($p < 0.001$).

Discussion

The goal of this study was to apply the RQA methodology to characterize the spatiotemporal properties of fixation patterns of 6- to 7-month-old infants during the face pop-out task. We employed a recently developed recurrence quantification analysis (RQA) to capture the temporal characteristics of scanpaths, its first use in this context, and compare them with the content of a scene (Anderson et al., 2013; Wu et al., 2014). We showed that the temporal dynamics of scanning varied systematically depending on the scene content (i.e., the presence of a face) and that the RQA measures provide more information about the temporal structuring of fixations than traditional measures of

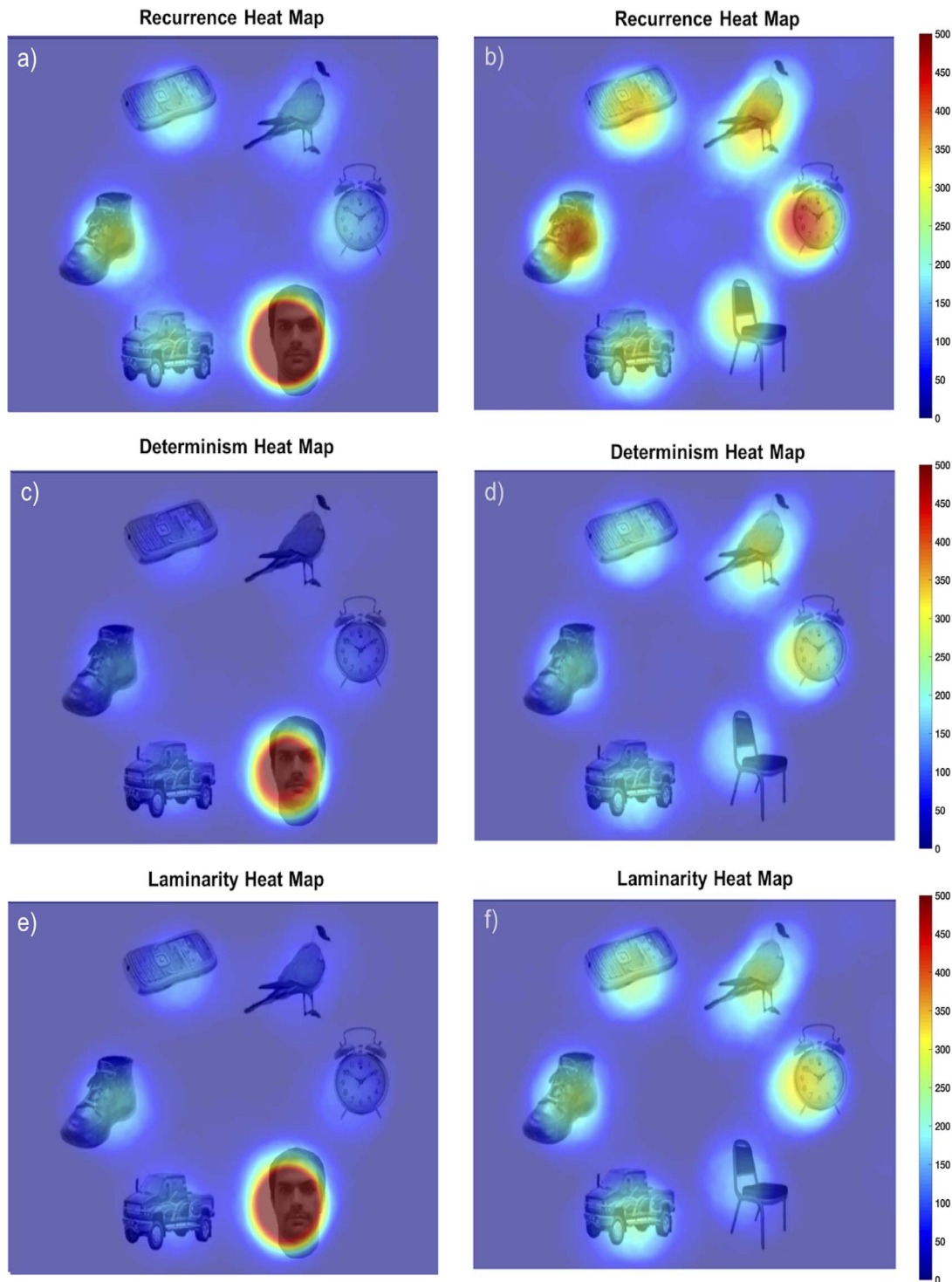


Figure 5. Example of the total heat maps obtained for one trial with a visual scene with a face (a, c, and e) and visual scene with a chair (b, d, and f). Recurrence rate heat maps, with only those fixations that were recurrent are represented in the image (a-b). Determinism heat maps, with only deterministic fixations are represented in the image (c-d). Laminarity heat maps, with only laminar fixations are represented in the image (e-f). The x- and y-axis corresponds to the horizontal and vertical dimension of the stimulus, respectively, and the color represents the amount of gaze data aggregated at each location within the stimulus. The scale on the right-hand side is applied to the two images located beside it. The human face is an example of what the original stimulus looks like. Written consent was given by the first author of this publication to use the image.

	Percentage of recurrent fixations (%)					
	Faces	Clocks	Cars	Birds	Mobiles	Shoes
Recurrence rate heat map	51.8	13.5	8.6	6.9	10.1	9.0
Determinism heat map	57.9	12.2	7.9	5.8	8.5	7.6
Laminarity heat map	62.5	12.7	6.1	4.8	7.6	6.1

Table 2. Percentage of recurrent fixations for each object within the visual scenes with faces for the recurrence rate, laminarity, and determinism heat maps.

accumulated looking (total fixation durations). Next, we showed that by normalizing the RQA measures with the information given by average fixation durations a more detailed account of the dynamics of fixation sequences can be obtained. Furthermore, we presented how these temporal dynamics were linked to the spatial location in the image, without previous definition of any areas of interest (AOIs), in order to explain the differences in scanning between visual scenes. Finally, we computed measures of stability of scanning and showed how the infants’ foraging strategies are driven by the image content.

RQA parameters on fixation-filtered data

Measures of scanning are divided into global (i.e., the distribution of fixation sequences over the whole image) and local (i.e., repeating sequences of fixations and detailed inspections of a particular image area; Anderson et al., 2013). RQA revealed significant differences in the global fixation patterns between visual scenes with faces and chairs. The RR was significantly higher when faces were present in the scene: the infants were more likely to refixate previously examined areas when a face was present. The CORM also showed significant differences in the distribution of refixation lags between both types of scenes. Higher CORM for face scenes indicated a longer temporal gap

	Percentage of recurrent fixations (%)					
	Chairs	Clocks	Cars	Birds	Mobiles	Shoes
Recurrence rate heat map	11.0	27.0	16.7	15.0	14.7	15.4
Determinism heat map	10.3	29.5	16.8	15.2	13.9	14.3
Laminarity heat map	9.2	31.0	17.4	13.9	14.9	13.4

Table 3. Percentage of recurrent fixations for each object within the visual scenes with chairs for the recurrence rate, laminarity and determinism heat maps.

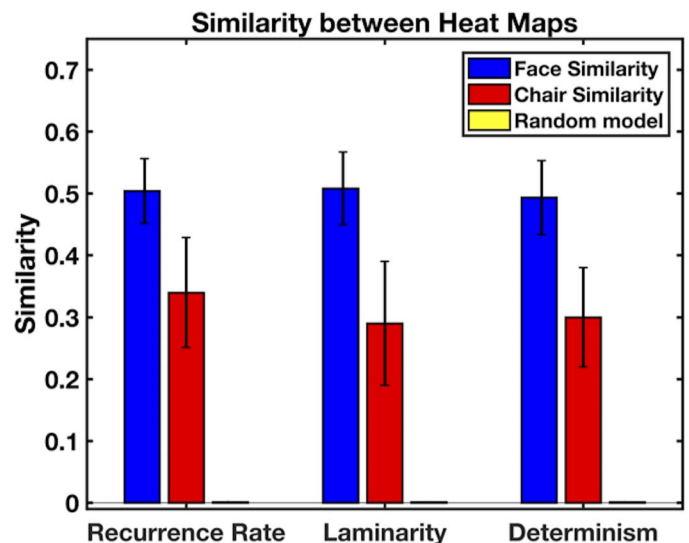


Figure 6. Average similarity values between the recurrence rate, laminarity, and determinism heat maps for the chairs (red) and faces (blue) conditions. Comparison between conditions showed highly significant differences for both types of scenes ($***p < 0.001$). In all cases, face and chair conditions were highly significant in comparison to the random model ($p < 0.001$). Error bars indicate the standard deviation.

between fixations and refixations, while chair scenes the CORM was lower, showing that refixations happen closer together in the fixation sequence. This decrease in the visual scenes with chairs could represent an adaptive response in which infants are less exogenously driven and are actively exploring the whole stimuli because no single highly salient stimulus (a face) is present (e.g., Gliga et al., 2009). It could be argued that the decreased CORM is due to a decrease in the total number of fixations in the visual scenes with chairs or to a decrease in the number of recurrences in them. In this case, the CORM would be much harder to interpret with varying levels of recurrence. However, the number of fixations in chair arrays was slightly higher than in the faces ones and the RR did not correlate with the CORM (see Supplementary File S1), which ruled out these effects.

RQA also showed differences in the local temporal gaze patterns. We found higher LAM values in the face versus chair scenes, which indicated that infants inspected particular areas in greater detail when a face was present. The DET was also significantly higher for the face scenes compared to the chair ones. This is related to the higher values of the RR and it means that the infants’ specific scanning patterns repeated more often. These patterns could be of at least two kinds: either a repeated sequence of fixations on consecutive objects (see black ellipse on Figure 2) or infants repeatedly had several consecutive fixations on the same objects (see blue square in Figure 2, whose

diagonal lines contribute also to the DET). The latter seems the most probable scenario in the face scenes judging from the increased LAM values. This might boost the DET and hide greater differences in scanning between participants. We believe that further analysis of this issue is required to find a parameter that eliminates the redundancy introduced by the laminarity measures and improves the computation of the DET independently of recurrent consecutive fixations on the same object for ET data.

Total fixation durations versus RQA parameters

We tested whether the RQA measures are related to the information provided by traditional measures of infant looking. First, we compared them with total fixation duration, which is a cumulative measure of attention affected by both endogenous processes and exogenous factors, such as stimulus properties (Wass & Smith, 2014). We observed that the total fixation duration for the face AOI was strongly positively correlated with the RR, the LAM, and the DET, explaining between 18% and nearly 45% of variance in these measures. This is unsurprising since in the pop-out task, faces are strong exogenous cues for attention in infancy (e.g., Gliga et al., 2009), and they not only capture attention but also hold it (Langton et al., 2008). If this is the case, not only would total fixation duration would be longer, but also a higher number of fixations related to the face would be expected, resulting in a higher number of revisitations (i.e., higher RR), more detailed scanning (i.e., higher LAM), or higher number of repeated fixation patterns (i.e., higher DET). This interpretation is consistent with significant, albeit lower, correlations of RQA measures with total fixation duration at the clock AOI, which also has face-like properties (Hadjikhani, Kveraga, Naik, & Ahlfors, 2009).

However, the information provided by the RQA and total fixation duration is essentially different. Total durations provide cumulative information about fixations, which has helped in the past to shed light on the perceptual and cognitive abilities of infants (e.g., Wass & Smith, 2014). However, such high-level attentional measures do not supply information about the temporal structuring of fixation sequences. In contrast, RQA provides information on the variability of scanning, revisitations over time (CORM), repetitive fixation structures (LAM), or even individual scanning patterns (DET). This allows for the comparison of fixation sequences of all the infants, and the subsequent use of these measures to flag different scanning patterns among them.

Average fixation duration versus RQA parameters

We also found lower, but significant, correlations between some RQA measures and average fixation durations for both face and chair slides. Individual differences in average FDs likely represent differences in several endogenous processes involving, for example, arousal or level of vigilance, as well as higher-order cognitive processes (Henderson & Pierce, 2008; Papa-georgiou et al., 2014; Wass et al., 2015). For this reason, in subsequent analyses, we normalized RQA values to control for differences in FDs. However, available studies show that stimulus properties and their capacity for exogenous cueing may also affect average FDs (Wass & Smith, 2014), thus higher correlations of the RR and the LAM with average FDs for the face than the chair slides may also reflect the differences in these stimuli.

Importantly, the CORM did not correlate with fixation duration measures. This parameter quantifies the temporal distribution of recurrences in fixation sequences and, to our knowledge, a similar measure has not been systematically investigated in infant eye-tracking. However, revisitations in visual scanning may serve as an early marker of atypical attention development in infants at risk of autism. Gliga, Smith, Gilhooly, Charman, and Johnson (2015) calculated the probability of revisitations by coding up to 10 visits (i.e., the sum of all consecutive fixations within an AOI). However, information about local temporal gaze patterns (i.e., LAM) gets lost by collapsing consecutive fixations into one visit. Furthermore, information about the temporal dynamics might also be lost since not all the fixations are included in the analysis. In our study, the CORM did not discard any information and included all the fixations in the sequence, offering a better description of the temporal distribution of refixation lags. We believe that the CORM is a new parameter that can provide more insight into the scanning strategies of infants in more complex stimuli (e.g., in dynamical videos) and for more complex settings such as visual scenes during real-life social interactions.

Normalized RQA

We showed that the normalization of the RQA measures with average FDs increased the existing differences in the RR, the LAM, and the CORM between the face and chair scenes. The normalized RR increased for the face scenes, indicating longer revisitations in the face scenes than the chair ones. Related to this, the LAM had a large decrease in the chair scenes, which suggests that consecutive fixations in

specific areas of the scene are shorter in the visual scenes with chairs than those with faces. Since these two parameters had a strong correlation with total fixation duration on the face AOI, the changes are consistent with previous studies reporting faces to hold attention longer in comparison to other objects (e.g., Langton et al., 2008). The variation of the CORM could follow a similar explanation. If refixations happening at later lags are related to a face being present, then the normalized CORM should be higher in comparison to the visual scenes with chairs. Altogether, the normalization process shows how high-level attentional measures affect the RQA measures and how interpreting the effects of normalization may help to understand in more detail the spatiotemporal properties of fixation sequences (see also Anderson et al., 2013; Vaidyanathan, Pelz, Alm, Shi, & Haake, 2014).

RQA heat maps

RQA heat maps were created by back-projecting the values of RR, DET, and LAM into image space to link them to their spatial location (Anderson et al., 2013). The maps confirmed that in the face scenes, the predominant object was, as expected, the face. Moreover, the LAM heat map showed that detailed scanning focused more on faces in comparison to the remaining objects. Interestingly, although in the visual scenes with chairs the distribution of recurrences was more even, clocks arose as the predominant object. This concurs with the aforementioned correlations and it could be related to clocks attracting more attention due to structural similarity with faces (Hadjikhani et al., 2009).

Next, we observed that the strong exogenous effect of led to an increase in the similarity of gaze patterns among infants for all the face scenes in comparison with chair scenes. In other words, when a face is present, it attracts attention and increases the similarity of scanning patterns between subjects. This is in contrast to the lower similarity for the chair scenes, where in the absence of strong exogenous cues there was larger between-subjects variability in scanning and lower similarity of gaze patterns. The laminarity and the determinism heat maps confirmed these effects and showed increased similarity in the face versus chair scenes. These results might have important implications for the quantification of individual differences in scanning, useful for studies of atypical social attention.

However, the similarity in scanning could be the result of the shared underlying structure of visual scenes, for example, the circular arrangement of objects. This may increase the chance that different scanning strategies return similar heat maps (i.e., false

positives). Although our task conditions lead to large face versus chair scene differences, we also used a more robust control for this effect by calculating similarity to a random model. Thus, future studies using these approaches should be cautious and consider the underlying structure of stimuli when comparing these heat maps in order to avoid false positives.

A clear advantage of the RQA heat map analysis over scanpath or FD methods is that it was carried out without the definition of AOIs. In many cases, this limits the analysis as potentially nearby fixations may be classified into spatially separated (Anderson et al., 2015). Also, AOIs that differ in size and shape might be problematic for several reasons (see Hessels et al., 2016a). RQA overcomes these problems and therefore provides a unique tool that in conjunction with heat maps can deal with the spatiotemporal analysis of fixation data (Anderson et al., 2013).

Limitations

We note several points that require further investigation or are limitations of RQA-based methods. First, the effects of changes to the radius size, which is the most critical parameter in RQA in fixation-filtered data. The radius represents the distance in which two fixations are considered recurrent. A large radius could lead to higher number of recurring time points and merge spatially close but distinct fixations. On the other hand, a small radius size might not be as robust in very noisy data as other methods (Hessels et al., 2016b). In this paper, we did not systematically test the effect of the radius size and 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). Additionally, as the data was recorded outside the lab and it is likely to be noisier, a relatively large radius proved to be sufficient for this analysis. However, checking the effect of radius sizes on the RQA measures can provide supplementary details about the viewing strategies of the participants (Wu et al., 2014). Additional work is needed to fully understand the effects of varying the radius size, and therefore future studies on this topic are required to establish its effects on the RQA parameters in relation to image content and data quality.

A drawback of using the current approach is the requirement of preprocessing to extract all the fixations. Although there are reliable automatic and semi-automatic tools for feature extraction (e.g., I2MC, Graphix), the poorer quality of infant relative to adult ET data means that none of these methods are fully reliable. A possible solution is to remove the preprocessing step and to use raw ET data, as demonstrated in

a recent study (Demiralp et al., 2017). However, infant data is more challenging due to its lower quality, so studies of RQA with raw (i.e., not fixation-filtered) infant ET data are necessary.

Future directions

In this paper, we introduced the application of the RQA to infant eye-tracking data. However, further work should provide an in-depth analysis of factors that drive infant scanning over time. One possibility involves measuring, how scanning gradually evolves during the task, which is something that cannot be studied with traditional high-level measures. This could be achieved using a moving window across the diagonal line of the recurrence plot in search for transitions in the RR and the LAM to identify the areas of the visual scene that attract attention and lead to more consecutive fixations suddenly falling in the same location. RQA could also be applied to identify transitions between fast and broad scanning by looking at the recurrences of durations of individual fixations. During exploration, fixations tend to be shorter, while when we focus our attention on particular area, FDs tend to be longer (e.g., Henderson & Pierce, 2008). Thus, these transitions could be explored by first applying RQA on FDs and then applying again a window through the diagonal line to identify transitions in the RR or LAM. Consequently, if specific patterns of scanning exist among participants (i.e., particular fixation sequences), it should be investigated if they simply reflect a hierarchy of low-level saliency, or actually a strategy of visual information sampling for specific classes of objects (e.g., faces, natural scenes, tools).

RQA could prove useful for studying visual scanning in atypical development. Children with autism spectrum disorder (ASD) demonstrate high revisitation rates when exploring natural visual scenes (Gliga et al., 2015). Additionally, infant siblings of children with ASD exhibit socio-communicative problems, which could be linked to difficulties with attention disengagement from socially relevant stimuli (Elsabbagh et al., 2013). Moreover, disorders such as ADHD are associated with novelty seeking and higher exploration in visual scenes (Gliga et al., 2015). As we have shown in this paper, the RQA measures are sensitive to scene type and content and thus, may prove useful in studying how preference for local over global features or low-level attention difficulties contribute to atypical scanning patterns of complex scenes with social stimuli (see Cheung, Bedford, Johnson, Charman, & Gliga, 2018). Therefore, this approach might capture differences in temporal patterns of visual foraging in infants and possibly serve as a tool for early detection atypical development of attention.

Conclusion

Our study sought to apply the RQA methodology, for the first time, to infant eye tracking data from a face pop-out task. We showed that the RQA measures are sensitive to the presence of faces during the face pop-out, and that they provide an in-depth description of the temporal dynamics of fixations in comparison to cumulative measures of looking. Moreover, RQA allows to normalize these measures with the information present within FDs, thus offering a more complete interpretation of the temporal dynamics of fixation sequences. We also demonstrated that the RQA heat maps provide a link between the temporal dynamics and the spatial properties of the image without the requirement of drawing any area of interest. Taken together, we believe that the RQA methodology provides a powerful tool for the study of fixation sequences in infancy research.

Keywords: recurrence quantification analysis (RQA), visual attention, eye-tracking, infants

Acknowledgments

Data collection was supported by the Nuffield Foundation (PI: DM). 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. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org.

We would like to 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 would especially like to thank Sally Parkinson, Head of Commissioning in Newham, for helping with setting up partnerships with Children's Centres, and Monica Forty and her team for the ongoing support in Children's services in Tower Hamlets. We also would like to thank Dr. Katie Crowley and Edyta Stanaszek for comments in the manuscript.

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 analyses was developed thanks to the Polish

National Science Centre Grant No. 2012/07/B/HS6/01464, awarded to PT.

Commercial relationships: none.

Corresponding authors: David López Pérez;

Przemysław Tomalski.

Email: david.lopez@psych.uw.edu.pl; p.tomalski@uw.edu.pl.

Address: Faculty of Psychology, University of Warsaw, Warsaw, Poland.

References

- Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior Research Methods*, *47*(4), 1377–1392, <https://doi.org/10.3758/s13428-014-0550-3>.
- 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.
- Aslin, R. N. (2012). Infant eyes: A window on cognitive development. *Infancy*, *17*(1), 126–140, <https://doi.org/10.1111/j.1532-7078.2011.00097.x>.
- Aslin, R. N. (2013). Infant learning: Historical, conceptual, and methodological challenges. *Infancy*, *19*(1), 2–27, <https://doi.org/10.1111/infa.12036>.
- Ballieux, H., Tomalski, P., Kushnerenko, E., Johnson, M. H., Karmiloff-Smith, A., & Moore, D. (2016). Feasibility of undertaking off-site infant eye-tracking assessments of neuro-cognitive functioning in early-intervention centres. *Infant and Child Development*, *25*, 95–111.
- Cherubini, M., Nüssli, M.-A., & Dillenbourg, P. (2010). This is it! Indicating and looking in collaborative work at distance. *Journal of Eye Movement Research*, *3*(3), 1–20.
- Cheung, C., Bedford, R., Johnson, M., Charman, T., & Gliga, T. (2018). Visual search performance in infants associates with later ASD diagnosis. *Developmental Cognitive Neuroscience*, *29*, 4–10, <https://doi.org/10.1016/j.dcn.2016.09.003>.
- Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225–234, <https://doi.org/10.1016/j.nlm.2008.06.002>.
- Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700, <https://doi.org/10.3758/brm.42.3.692>.
- Demiralp Ç., Cirimele J., Heer J., & Card S. K. (2017). The VERP Explorer: A tool for exploring eye movements of visual-cognitive tasks using recurrence plots. In M. Burch, L. Chuang, B. Fisher, A. Schmidt, & D. Weiskopf (Eds.), *Eye tracking and visualization: Foundations, techniques, and applications* (pp. 41–55). *ETVIS 2015. Mathematics and Visualization*. New York: Springer. https://doi.org/10.1007/978-3-319-47024-5_3.
- Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior Research Methods*, *44*(4), 1079–1100, <https://doi.org/10.3758/s13428-012-0212-2>.
- Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., & Johnson, M. H. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, *74*(3), 189–194, <https://doi.org/10.1016/j.biopsych.2012.11.030>.
- Fantz, R. L. (1963). Pattern vision in newborn infants. *Science*, *140*(3564), 296–297, <https://doi.org/10.1126/science.140.3564.296>.
- Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670.
- Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: An experimental test of scanpath theory. *Journal of Experimental Psychology: General*, *142*(1), 41–56, <https://doi.org/10.1037/a0028227>.
- Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170, <https://doi.org/10.1016/j.cognition.2008.11.010>.
- Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*, 550–562, <https://doi.org/10.1080/15250000903144199>.
- Gliga, T., Smith, T., Gilhooly, N., Charman, T., & Johnson, M. H. (2015). Early visual foraging in relationship to familial risk for autism and hyperactivity/inattention: A preliminary study. *Journal of Attention Disorders*, *22*(9), 839–847, <https://doi.org/10.1177/1087054715616490>.
- Großekathöfer, U., Manyakov, N. V., Mihajlović, V., Pandina, G., Skalkin, A., Ness, S., & Goodwin, M. S. (2017). Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. *Frontiers in Neuroinformatics*, *11*: 9.

- Hadjikhani, N., Kveraga, K., Naik, P. & Ahlfors, S. P. (2009). Early (M170) activation of face-specific cortex by face-like objects. *Neuroreport*, *20*, 403–407, <https://doi.org/10.1097/WNR.0b013e328325a8e1>.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. (2007). Image salience versus cognitive control of eye movements in real-world scenes: Evidence from visual search. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 537–562). Oxford, UK: Elsevier.
- Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, *15*(3), 566–573, <https://doi.org/10.3758/pbr.15.3.566>.
- 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., Kemner, C., van den Boomen, C., & Hooge, I. T. C. (2016a). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods*, *48*(4), 1694–1712.
- Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. (2016b). 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>.
- Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 45–52). New York: ACM.
- 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\(91\)90045-6](https://doi.org/10.1016/0010-0277(91)90045-6).
- Johnson, M. H., & de Haan, M. D. (2015). *Developmental cognitive neuroscience: An introduction*. Hoboken, NJ: Wiley-Blackwell.
- Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*(4), 335–344.
- 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>.
- Kato, M., & Konishi, Y. (2013). Where and how infants look: The development of scan paths and fixations in face perception. *Infant Behavior and Development*, *36*(1), 32–41, <https://doi.org/10.1016/j.infbeh.2012.10.005>.
- Kennedy, D. P., D'Onofrio, B. M., Quinn, P. D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic influence on eye movements to complex scenes at short timescales. *Current Biology*, *27*(22), 3554–3560, <https://doi.org/10.1016/j.cub.2017.10.007>.
- Langton, S. R., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*(1), 330–342, <https://doi.org/10.1016/j.cognition.2007.07.012>.
- Le Meur, O. L., & Baccino, T. (2012). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior Research Methods*, *45*(1), 251–266, <https://doi.org/10.3758/s13428-012-0226-9>.
- López Pérez, D., Leonardi, G., Niedźwiecka, A., Radkowska, A., Raczaszek-Leonardi, J., & Tomalski, P. (2017). Combining recurrence analysis and automatic movement extraction from video recordings to study behavioral coupling in face-to-face parent–child interactions. *Frontiers in Psychology*, *8*:2228, <https://doi.org/10.3389/fpsyg.2017.02228>.
- Nomikou, I., Leonardi, G., Rohlfing, K., & Raczaszek-Leonardi, J. (2016). Constructing interaction: The development of gaze dynamics. *Infant and Child Development*, *25*(3), 277–295, <https://doi.org/10.1002/icd.1975>.
- Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychological Science*, *25*(7), 1371–1379, <https://doi.org/10.1177/0956797614531295>.
- Richards, J. E. (2008). Attention in young infants: A developmental psychophysiological perspective. In C. A. Nelson & M. Luciana (Eds.), *The handbook of developmental cognitive neuroscience* (2nd ed., pp. 479–498). Cambridge, MA: MIT Press.
- Richardson, D., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*, 39–54.

- Romero, V., Fitzpatrick, P., Schmidt, R., & Richardson, M. J. (2016). Using cross-recurrence quantification analysis to understand social motor coordination in children with autism spectrum disorder. In C. L. Webber, Jr., C. Ioana, & N. Marwan (Eds.), *Recurrence plots and their quantifications: Expanding horizons* (pp. 227–240), Springer Proceedings in Physics, Vol. 180. New York: Springer International Publishing
- Saez de Urabain, I. R., Nuthmann, A., Johnson, M. H., & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research* 134, 43–59.
- Shockley, K., Santana, M., & Fowler, C. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–332.
- Shockley, K., & Turvey, M. (2005). Encoding and retrieval during bimanual rhythmic coordination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 980–990.
- Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2012). Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. *Infancy*, 18(4), 534–553, <https://doi.org/10.1111/j.1532-7078.2012.00135.x>.
- Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E. L., Murphy, E., Moore, D. G., & Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology*, 10(5), 611–624, <http://doi.org/10.1080/17405629.2012.728076>.
- Vaidyanathan, P., Pelz, J., Alm, C., Shi, P., & Haake, A. (2014). Recurrence quantification analysis reveals eye-movement behavior differences between experts and novices. *Proceedings of the Symposium on Eye Tracking Research and Applications—ETRA '14* (pp. 303–306), New York: ACM. <https://doi.org/10.1145/2578153.2578207>.
- Vallee, J. (2015). Un cadre d'apprentissage automatique appliqué à des données oculométriques décontextualisées [A Machine Learning Framework Applied to Decontextualized Eye Tracking Data]. (Unpublished master's thesis). HEC Montréal, Montréal, Quebec, Canada.
- Wass, S. V., Jones, E. J., Gliga, T., Smith, T. J., Charman, T., & Johnson, M. H. (2015). Shorter spontaneous fixation durations in infants with later emerging autism. *Scientific Reports*, 5(1), <https://doi.org/10.1038/srep08284>.
- 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>.
- Wass, S. V., Smith, T. J., & Johnson, M. H. (2012). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, 45(1), 229–250, <https://doi.org/10.3758/s13428-012-0245-6>.
- 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, 1–14, <https://doi.org/10.1167/14.9.8>. [PubMed] [Article]
- Xiao, W. S., Xiao, N. G., Quinn, P. C., Anzures, G., & Lee, K. (2012). Development of face scanning for own- and other-race faces in infancy. *International Journal of Behavioral Development*, 37(2), 100–105, <https://doi.org/10.1177/0165025412467584>.
- Zbilut, J. P., Giuliani, A., & Webber, C. L., Jr. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters. A*, 246, 122–128.