

Modeling gaze behavior with continuous-time markov chains to investigate social attention dynamics in autism

Article

Published Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Open Access

Bruschetta, R., Famà, F. I., Spadaro, L., Leonardi, E., Carrozza, C., Aiello, S., Campisi, A., Mastrogiuseppe, M., Campisi, S., Chakrabarti, B. ORCID: <https://orcid.org/0000-0002-6649-7895>, Ruta, L., Pioggia, G., Gaetano, A. D., Borri, A., Tartarisco, G., Boncoddò, M., Crimi, I., Bella, F. D., Lazzaro, G., Martines, S., Orsina, A. and Rosano, M. (2025) Modeling gaze behavior with continuous-time markov chains to investigate social attention dynamics in autism. Scientific Reports, 15. 39692. ISSN 2045-2322 doi: 10.1038/s41598-025-23366-4 Available at <https://centaur.reading.ac.uk/127170/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1038/s41598-025-23366-4>

Publisher: Nature Publishing Group

including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



OPEN Modeling gaze behavior with continuous-time markov chains to investigate social attention dynamics in autism

R. Bruschetta¹, F. I. Famà¹, L. Spadaro¹, E. Leonardi¹, C. Carrozza¹, S. Aiello¹, A. Campisi^{1,2}, M. Mastrogiuseppe^{1,2}, S. Campisi¹, B. Chakrabarti³, L. Ruta^{1✉}, G. Pioggia^{1✉}, A. De Gaetano^{1,4,5,6}, A. Borri⁴, G. Tartarisco¹ & The NEST Team^{*,*}

This study introduces a novel methodological framework combining continuous-time Markov chains and principal component analysis (PCA) to model and investigate gaze behavior in young children observing naturalistic social interactions. By quantifying transition propensities between areas of interest (AOIs), this approach enables a dynamic, data-driven analysis of gaze patterns beyond static fixation metrics. We applied this framework to eye-tracking data from children with autism spectrum condition (ASC) and neurotypical (NT) peers as they watched scenes of a child and an adult engaged in interactive play, involving turn-taking and reciprocal imitation. The stimuli, designed to ensure ecological validity, depicted sensory social routines (SSRs) with songs and shared play with musical instruments, allowing exploration of gaze dynamics in both dyadic and triadic social contexts. Results revealed distinct gaze transition profiles in ASC children, characterized by more frequent disengagement from socially salient AOIs and reduced re-orientation to faces following non-social fixations. In contrast, NT children exhibited frequent gaze alternation between faces and triangulation with objects, supporting joint attention and reciprocal engagement. Additionally, ASC participants were more likely to enter and persist in non-social states, especially during object-centered trials, whereas NT peers showed consistent transitions toward socially meaningful targets. These findings highlight the relevance of capturing the temporal patterns of visual engagement in autism, revealing how moment-to-moment gaze transition dynamics reflect underlying differences in social motivation, attentional control, and sensory processing. The proposed framework provides a powerful tool for characterizing individual differences in gaze organization and holds promise for advancing biomarker identification in neurodevelopmental research.

Keywords Eye-tracking, Autism, Gaze patterns, Social attention, Markov chains, Principal component analysis

Reduced visual orientation toward social stimuli, such as gaze direction, facial expressions and gestures, typically referred to as social attention, has been consistently reported in autistic children from the earliest stages of development. These early differences in attentional engagement with socially salient cues can influence how children learn from social interactions and shape their social cognitive developmental trajectories^{1–4}.

To investigate social attention in autism, eye-tracking paradigms have traditionally employed a variety of stimuli, ranging from static images^{5–8} to dynamic videos^{9–14}. Among these, dynamic and naturalistic paradigms have proven more sensitive in capturing reduced orientation to social cues in autistic populations^{15–18}, as they more closely mirror real-world social contexts. Such stimuli require children to process complex multimodal

¹National Research Council of Italy, IRIB, Istituto Marino, Via Leanza, 98164 Mortelle, Messina, Italy. ²University of Trieste, Trieste, Italy. ³Centre for Autism, School of Psychology & Clinical Language Sciences, University of Reading, Reading, UK. ⁴National Research Council of Italy, IASI, Rome, Italy. ⁵Department of Biomaterials, Óbuda University, Bécsi Road 96/B, H-1034 Budapest, Hungary. ⁶Department of Mathematics, Mahidol University, Bangkok, Thailand. *A list of authors and their affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper. ✉email: liliana.ruta@irib.cnr.it; giovanni.pioggia@irib.cnr.it

signals, integrating visual, auditory, and motor information, and flexibly shifting attention between social and non-social elements in real time.

While most eye-tracking studies rely on screen-based paradigms, complementary evidence from live interaction settings confirms similar atypicalities in social gaze. For example, face-to-face paradigms have shown that autistic children allocate less attention to partners' faces and gestures during naturalistic exchanges^{19,20}.

Despite this progress, most studies still rely on static measures of gaze allocation within specific areas of interest (AOIs), providing only a partial picture of how attention unfolds over time. The temporal gaze structure instead captures the dynamic organization and sequencing of gaze shifts, considering not only the duration of attention within specific AOIs but also the order, frequency, and timing of transitions between them. By moving beyond static dwell-time metrics, this dynamic perspective provides a richer characterization of attentional engagement, enabling the identification of subtle and fine-grained patterns that may better reflect the atypical dynamics of social attention often observed in autism.

Although a few studies have begun to explore these temporal aspects, reporting patterns such as a gradual decline in face-directed gaze²¹ or diminished re-engagement with faces over time²², little is still known about how gaze transitions between social and non-social elements evolve dynamically in interactive contexts.

Over the past two decades, Markov-based approaches have gained increasing relevance for modeling these sequential aspects of visual attention. In particular, Hidden Markov Models (HMMs) have been widely used to represent gaze behavior as a probabilistic process unfolding across latent states, with transitions capturing the likelihood of shifting between regions of interest and emissions describing the spatial variability of fixations^{23–25}. These methods have been applied to diverse contexts, from face processing^{24,25}, to tracking of moving objects²⁶, visual inspection tasks²⁷, user classification²⁸, and human–computer interfaces²⁹. However, conventional HMM formulations rely on discrete-time representations and capture fixation duration only indirectly through sampling frequency or self-transitions, limiting their ability to fully characterize the fine-grained temporal dynamics of attention. Duration-aware variants such as hidden semi-Markov models (HSMMs) attempt to address this limitation but often at the expense of interpretability and analytical tractability^{30,31}.

In this context, continuous-time Markov chain (CTMC) models offer a compelling alternative, as they represent gaze behavior as a continuous-time stochastic process governed by a rate matrix (Q) that simultaneously encodes state-specific transition rates and dwell times³⁰. Notably, Li and colleagues³² demonstrated the advantages of continuous-time formulations in modeling the temporal dynamics of cognitive processes during strategic decision-making tasks, showing that explicitly parameterizing dwell times and transition rates yields more accurate and interpretable representations of underlying behavioral states. This formulation enables a direct and transparent description of attentional dynamics, decoupling temporal resolution from sampling frequency and allowing the derivation of theoretically grounded descriptors such as stationary distributions, mean hitting times, and transient probabilities. Despite their potential, CTMC approaches remain underexplored in the study of social attention in autism, particularly in naturalistic, socially rich contexts.

To fill this gap, the present study introduces an analytical framework for modeling the fine-grained temporal dynamics of visual attention in autism. Specifically, a Continuous-Time Markov Chain (CTMC) approach is employed to capture how children transition between multiple AOIs over time while observing naturalistic dyadic interactions. This probabilistic modeling technique allows a formal description of the sequential structure of gaze patterns, revealing deeper attentional dynamics beyond what static metrics can provide. In addition, Principal Component Analysis (PCA) is applied to extract key dimensions of variability in gaze behavior, enabling a data-driven characterization of individual profiles.

These methods are implemented within an innovative eye-tracking paradigm using ecologically valid, video-recorded play scenarios between a child and an adult partner. The interactions include structured object-based activities with musical instruments (xylophone and drums) and gesture-accompanied nursery rhymes, creating a rich, multimodal setting that promotes shared attention, reciprocal turn-taking, and mutual imitation. The task design carefully controls critical parameters such as role alternation, sensory consistency, distractor placement, and activity structure to isolate meaningful differences in gaze behavior. By focusing on the temporal organization of gaze shifts and leveraging advanced computational tools, the study aims to provide novel insights into the dynamics of social attention in autism.

Based on prior evidence of reduced social attention in autism, we expected that:

1. Across social routines and object-based activities, autistic children would show fewer transitions between socially salient AOIs (faces) compared to neurotypical peers.
2. Autistic children would be more likely to maintain their gaze within non-social AOIs (activities or distractors), while neurotypical children would display more flexible gaze shifts integrating social and non-social elements.
3. These group-level differences would emerge more clearly when applying a dynamic modeling approach (CTMC) than when using conventional static measures of gaze allocation (dwell time).

Methods

Participants

The participants in this study were recruited as part of a larger project investigating visual attention and social communication in autistic and non-autistic children. For the present analysis, we specifically focused on the examination of gaze transitions across different areas of interest.

The sample consisted of 55 preschoolers, aged between 29 and 93 months, a developmental window in which social attention and gaze strategies undergo substantial refinement and atypicalities in autistic children are particularly evident^{33,34}. Of these, 24 children were clinically diagnosed with Autism Spectrum Condition (ASC), based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria³⁵.

Expert clinicians performed the diagnostic assessments, using the Autism Diagnostic Observation Schedule - Second Edition³⁶ as a supporting tool. These assessments took place at the Institute for Biomedical Research and Innovation of the National Research Council of Italy (IRIB-CNR) in Messina and at the Centre for Autism Spectrum Disorders, Child Psychiatry Unit, Provincial Health Service (ASP-CT) in Catania, Italy. The neurotypical (NT) group included 31 children, who were recruited from two mainstream nursery schools in Messina.

The final sample size reflects the pool of families who consented to participate and met inclusion/exclusion criteria during the recruitment period. The inclusion criteria for the ASC group required the absence of known genetic syndromes (e.g., fragile X syndrome, tuberous sclerosis), inborn metabolic disorders (e.g., aminoaciduria, peroxisomal disorders), epilepsy with uncontrolled seizures, movement disorders, or cerebral palsy. For the NT group, exclusion criteria included any clinical diagnosis of neurodevelopmental conditions (e.g., language and/or motor delays/disorder, intellectual disability, genetic syndromes, epilepsy) and a family history of autism. The two groups were matched for age, with no significant differences ($p > 0.05$).

All participants had normal or corrected-to-normal vision and no history of auditory impairment. The study was approved by the Ethics Committee of the National Research Council (CNR; ethical clearance, 01.08.2018) and by the local health authority ASP-CT (Prot. N. 498). All methods were carried out in accordance with relevant guidelines and regulations, and in compliance with the Declaration of Helsinki. Informed consent was obtained from all caregivers (i.e., the legal guardians) for their children's participation in the study.

Task Procedure

During the eye-tracking experiment, children sat in a small chair within a quiet, controlled environment, positioned 80 cm from a high-resolution 24" widescreen LCD monitor (1024 x 768 pixels). A research team member (S.L.) ensured the children's engagement and promptly repositioned them if they moved outside the trackable range. Gaze patterns were recorded with the SMI iView XTM RED dark-pupil 250 Hz eye-tracking system and exported using SMI BeGaze 2.4 software.

Before starting the task, a nine-point calibration with automatic validation was performed using dynamic, child-friendly targets, like a cartoon cat with meowing sounds, to maximize attention. Participants with a calibration error exceeding approximately 1.2° of visual angle were excluded from the analysis. Calibration and validation were repeated when necessary to ensure data quality³⁷.

As a result, three children (two ASC and one NT) were excluded due to calibration errors, while one ASC child did not complete the task. This resulted in a final analyzable sample of 51 children (22 ASC and 29 NT). Table 1 summarizes the demographic and clinical characteristics of the final sample.

Participants watched four 25-second videos in random order, each depicting a naturalistic interaction between a child and an adult seated at a small table (Fig 1). The interactions were recorded in a familiar playroom setting to maximize ecological validity, with real social partners engaging in structured but naturalistic activities. Two videos featured joint musical activities with instruments (*Drums* and *Xylophone*), while the other two depicted sensory social routines (SSRs) with songs and body gestures (*Sheriff* and *Witches*). In one musical activity (*Drums*) and one song routine (*Sheriff*), the child initiated the action, with the adult imitating after 2–3 seconds; in the other two videos (*Xylophone* and *Witches* respectively), the roles were reversed, with the adult initiating the sequence. Each clip included four sequential actions performed by each partner, incorporating reciprocal imitation and turn-taking. To ensure experimental control, several parameters were systematically balanced across videos, including: (1) interaction type (dyadic sensory social routines vs. triadic object-centered play), (2) role alternation (child-initiated vs. adult-initiated sequences), (3) sensory consistency across conditions, and (4) placement of distractor objects, positioned on a shelf at the left and right corners of the background. Each video began with a drift correction sequence followed by a 5-second black screen to standardize the start of the trials. This design allowed children to observe rich, multimodal social scenes closely resembling everyday interactions, while maintaining precise control over task parameters for comparability across conditions.

Data preprocessing

The SMI system was used to define seven distinct areas-of-interest (AOIs) for a detailed analysis of gaze patterns. These AOIs were: (1) Adult Face, (2) Child Face, (3) Adult Activity, (4) Child Activity, (5) Left Distractor Object,

Variable	ASC	NT	Statistic	p
N (female, %)	22 (4, 18.2%)	29 (19, 65.5%)	$\chi^2 = 9.49$	<0.01
Age (months) (mean ± SD)	56.64 ± 19.77	59.43 ± 8.84	$t = -0.68$	>0.05
Total DQ (mean ± SD)	76.53 ± 24.21	120.16 ± 14	$t = -8.10$	<0.0001
ADOS 2 – SA (mean ± SD)	9 ± 4.84	-	-	-
ADOS 2 – RRB (mean ± SD)	5.25 ± 2.67	-	-	-
ADOS 2 – Total (mean ± SD)	14.18 ± 7.40	-	-	-

Table 1. Demographic and clinical characteristics of the participants. *Note.* DQ = Developmental Quotient; SA = Social Affect; RRB = Restricted and Repetitive Behavior.

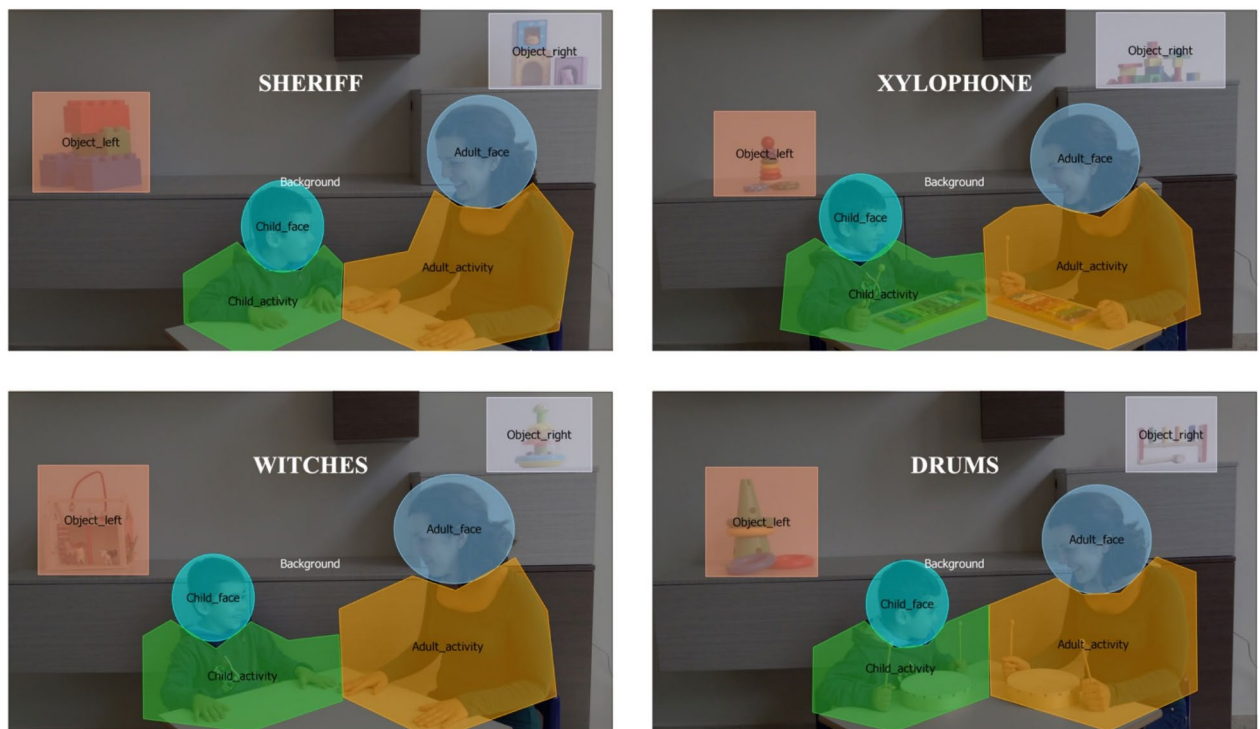


Fig. 1. Frame from each video clip with overlapped Areas of Interest (AOIs). The defined regions include: Adult Face, Child Face, Adult Activity, Child Activity, Left Object, Right Object, and Background.

(6) Right Distractor Object, (7) Background area. The Adult and Child Face AOIs primarily included the head and neck of each actor, with a small portion of the upper shoulders occasionally encompassed due to the shape of the bounding regions and natural movements during the interaction. The Adult and Child Activity AOIs covered the upper body, arms, hands, and any object being used in the object-based play videos (Fig 1). AOIs were manually adjusted each second to account for natural movement during interactions, ensuring consistent and precise data capture throughout the entire duration of each trial.

Trials with over 25% track loss were excluded from the eye-tracking analysis, resulting in the omission of 9 out of 204 trials (4.4%) (see Table S1, Supplementary Materials). To handle missing data, imputation was performed using the substitution technique with the median value.

Benchmark gaze measure (dwell time)

As a conventional and descriptive index of visual exploration, we computed dwell time, defined as the percentage of total looking time spent on each AOI. For this analysis, we considered seven AOIs as defined in the SMI system: Adult Face, Child Face, Adult Activity, Child Activity, Left Object, Right Object, and Background. Including the Background AOI allowed us to provide a comprehensive description of gaze allocation across the entire visual scene, in line with standard practice in eye-tracking studies.

It is important to note, however, that the Background AOI was excluded from the subsequent Markov chain analyses, which were restricted to six socially and functionally relevant AOIs (faces, activities, and distractor objects). This distinction reflects the different purposes of the two analyses: while dwell time offers a static and overall measure of attention distribution, the continuous-time Markov chain (CTMC) modeling captures the temporal organization of gaze transitions between socially and non-socially meaningful regions.

Markov chain data modeling

In this section, we outline the continuous-time Markov chains (CTMCs) used to model gaze patterns within the defined Areas of Interest (AOIs) as exhibited by children while watching the videos. A CTMC is a stochastic process characterized by a state space S , namely a finite or countable set representing all possible states, which in our case correspond to different AOIs—and by the propensities (or transition rates) that describe how quickly transitions occur between states per unit of time, providing a direct quantitative measure of the dynamics of gaze shifts. Unlike the more conventional discrete-time Markov chain (DTMC), which evolves in fixed time steps and is defined by transition probabilities, the CTMC framework allows transitions to occur at any continuous point in time. Although our experiment used a digital system with a theoretically fixed sampling interval, which might suggest that a DTMC approach could suffice, employing CTMC models offers an advantage. Specifically, they can account for uncertainties or slight delays in the data capture process, phenomena that we have indeed observed.

The evolution of a finite time-homogeneous CTMC (with n states) is characterized by a transition matrix Q , whose general element $q_{ij}()$ denotes the propensity of a transition from state i to j and q_{ii} represents the opposite of the probability of leaving the state i , which ensures that each column of Q sums to zero.

The dynamics of the probability $P(X_t = i)$ that the CTMC at a general time t is found in the general state i is governed by the following system of n ordinary differential equations:

$$\frac{d}{dt}P(X_t = i) = \sum_{j \neq i} P(X_t = j) \cdot q_{ji} - P(X_t = i) \cdot \sum_{j \neq i} q_{ij} \quad (1)$$

The above equation system, called Master Equation, captures the continuous-time evolution of the CTMC probability distribution, where transitions among states are probabilistic and occur with propensities specified by the elements of the transition matrix Q ³⁸.

In our modeling framework, each AOI is represented as a distinct state within the Markov chain, denoted as X_t at time t , indicating the specific AOI capturing the child's visual attention. This approach offers a sophisticated means to capture and analyze the nuanced transitions in gaze behavior over time, facilitating a detailed examination and prediction of gaze patterns across the defined AOIs. For each participant and each of the four trials, we computed a 6-state CTMC ($n=6$) using the following Areas of Interest (AOIs): Child Face, Adult Face, Child Activity, Adult Activity, Left Distractor Object, and Right Distractor Object. As a result, for each child and video stimulus, we obtained a 6×6 transition matrix, from which $n \cdot (n-1) = 30$ transition propensities were estimated for each trial, following the procedure outlined below.

According to the theory of continuous-time Markov processes, the expected waiting time in a state j before transitioning to state i is an exponential random variable with the mean equal to the inverse of the propensity from j to i . Therefore, for each child and trial, we estimated the general element $q_{ij}()$ of the transition matrix Q as the statistical mean of the observed waiting times from state j to state i . In cases where a state j was not visited for a given participant and trial, the corresponding column j of matrix Q would be a zero vector. This would lead to undesirable properties of the underlying behavior, particularly the non-uniqueness of the CTMC equilibrium distribution. To prevent such issues, when a state was not observed, we imputed the transition propensities from the non-visited state j by calculating the median value of the corresponding transition propensities from state j across all other subjects within the same group and trial. As a result, the unique equilibrium probabilities from the Master Equation were computed for each individual and trial, and these probabilities were analyzed as described in the next subsection.

Although the AOIs capture the main social and non-social elements of the scene, they are participant-centered and do not explicitly represent stimulus-level interaction events, such as moments of joint attention between the actors in the video (e.g., both attending to the same object). This design choice allowed us to focus on the sequential structure of participants' visual exploration. However, future extensions of the model could integrate event-based annotations of the stimulus to examine participants' alignment with interaction dynamics.

Consistent with the Markov property, the CTMC framework assumes that gaze transitions are "memoryless", meaning that the probability of moving to a new AOI depends only on the current state and not on the sequence of prior states or on the time already spent in that state. This simplifying assumption provides a tractable and interpretable representation of gaze dynamics, while effectively capturing the sequential organization of visual attention in naturalistic contexts.

Principal component analysis

To investigate whether neurotypical gaze patterns showed preferences for specific areas of the video stimuli, we first analyzed the equilibrium probabilities for the six AOIs across the four trials in the NT group. Specifically, we compared the median values of the equilibrium probabilities for each AOI using the non-parametric Kruskal-Wallis test, followed by pairwise comparisons with the Dwass-Steel-Critchlow-Fligner method. The analysis revealed that SSRs primarily drew NT children's attention to faces, whereas musical activities with instruments directed their gaze toward the activity areas. Specifically, in the *Sheriff* video stimulus, the Child Face AOI garnered significantly more attention than other areas, whereas in the *Witches* video, the Adult Face AOI was the primary focus. This pattern aligns with our expectations, as the *Sheriff* song is initiated by the child, and the *Witches* song by the adult. Conversely, during the *Drums* and *Xylophone* videos, the activity areas of both the child and adult were significantly more attended to than the facial areas (see Table S2 and Table S3, Supplementary Materials), consistent with object-based activities naturally drawing attention to materials and actions. Building on these findings, we focused subsequent analyses on the transitions most relevant to each type of task: face-related AOIs during the SSRs trials (*Sheriff* and *Witches*), and activity-related AOIs for the musical instrument trials (*Drums* and *Xylophone*).

This approach resulted in 18 distinct features for each trial, totaling 36 features for each stimulus group (SSRs and musical activities with instruments).

To handle this high-dimensional feature space and identify dominant patterns of gaze dynamics, we conducted a Principal Component Analysis (PCA), a standard technique for dimensionality reduction that projects data into a lower-dimensional space, while retaining as much variance as possible³⁹. This method allowed us to visualize group-level patterns and individual variability, revealing the distinctive structure of gaze exploration in NT and ASC children.

The PCA scatter plots for both video groups display participants projected onto the first two principal components, while the accompanying correlation plots illustrate how each feature contributes to these components, highlighting the transitions that drive the observed variability.

For clarity, we grouped features into five categories and applied a consistent color-coding scheme. In the SSRs trials, these groups included: (1) transition propensities between the Child Face and the Adult Face (and vice versa), shown in green, (2) transition propensities from the Adult Face to other areas excluding the Child Face, shown in blue, (3) transition propensities from the Child Face to other areas excluding the Adult Face, shown in red, (4) transition propensities towards the Adult Face from other areas excluding the Child Face, shown in pink, and (5) transition propensities towards the Child Face from other areas excluding the Adult Face, shown in gray. The same logic was applied to the musical instrument trials, where the reference AOIs were the child's and adult's activity areas. Details of these feature groupings are provided in Table S4 of the Supplementary Materials.

To formally test the visual group separation observed in the PCA scatter plots, we conducted a one-way MANOVA using the scores of the first two principal components as dependent variables and Group (ASC vs. NT) as the independent factor. Separate analyses were performed for the SSRs and musical activities.

Data visualization with chord diagrams

We included chord diagrams to visually represent the complex patterns of gaze transitions, providing a clearer and more intuitive understanding of the differences in gaze behavior between the ASC and NT groups. For transparency, the group-level average transition matrices used to generate these visualizations are provided in the Supplementary Materials (Tables S5–S12), one for each trial. Transition propensities between areas of interest (AOIs) are depicted through a simplified arrow-based design. Each AOI is represented as a segment along the circumference of the diagram, with directed transitions illustrated by curved arrows connecting the segments. The thickness of the arrows reflects the strength of the transition, enabling an immediate visual comparison of gaze shifts between AOIs. The arrow's direction indicates the direction of the gaze shift, moving from the starting AOI to the target AOI. The color of the arrow corresponds to the starting AOI and matches the color of the outer circle segment representing that AOI. At the base of the arrow, a colored segment represents the target AOI and matches the color of the corresponding inner circle segment. This design helps to visually distinguish the origin and destination of each transition. We generated these chord diagrams for each group (ASC and NT) and for each trial (*Sheriff*, *Witches*, *Drums*, and *Xylophone*). Self-transitions (i.e., transitions where the gaze remains within the same AOI) were excluded to reduce visual clutter, as their high propensity could dominate the visualization. Additionally, we applied a scaled version of the chord diagram, in which all AOI segments on the circumference were set to have equal size. Within each segment, however, the arrows representing transitions were proportionally scaled to reflect the fraction of interactions directed toward other AOIs. This scaling approach normalized the sector sizes while preserving the relative strength of transitions. In this way, the chord diagrams summarize the most frequent transitions between AOIs and clearly represent the typical gaze exploration paths for each group across conditions, capturing the multivariate complexity of gaze dynamics in a compact and accessible format.

Results

Markov-chain analysis of gaze dynamics

Using continuous-time Markov chains (CTMCs) to model gaze transition patterns across defined AOIs, we observed distinct group-level differences during both the SSRs involving songs and the musical activities with instruments.

Specifically, in the *Sheriff* and *Witches* trials, NT children showed significantly higher transition propensities between face-related AOIs (i.e., Adult Face and Child Face) compared to ASC children. This increased frequency of gaze shifts between faces is visually represented by the green arrows in the PCA space, where the average cluster vectors highlight stronger bidirectional transitions between faces in the NT group (Fig 2). The full set of loadings, including the contribution of each individual feature, is provided in the Supplementary Materials (Figure S1). Consistently, the chord diagrams further highlight these differences, showing thicker connections between Adult Face and Child Face for NT children (Fig 3). In particular, the arrow from Adult Face to Child Face is larger for NT children in the *Sheriff* trial, and the arrow from Child Face to Adult Face is more pronounced for NT children in the *Witches* trial.

Moreover, NT children also exhibited a significantly higher propensity to reorient attention from distractor objects or activity areas toward faces, as shown by the pink and gray arrows in the PCA visualization (Fig 2; see also the full feature-level representation in Figure S1). This tendency is further confirmed by the chord diagrams (Fig 3), where the light gray arrow from Object Left to Child Face, the dark gray arrow from Object Right to Adult Face, and the red arrow from Adult Activity to Adult Face are all larger in NT children compared to the ASC group. These patterns indicate a stronger inclination to maintain attention on social elements, as seen in the frequent gaze shifts between faces, and to reorient attention toward faces when initially directed toward non-social elements during visual exploration.

In contrast, ASC children demonstrated a higher propensity to avert their gaze from faces, directing their attention instead to non-social elements in the scene, such as distractor objects and activity areas, as indicated by the red and blue arrows in the PCA visualization (Fig 2 and S1). This pattern is further evident in the chord diagrams, where the green arrows from Child Face to Child Activity and from Child Face to Object Left, as well as the blue arrows from Adult Face to Adult Activity and from Adult Face to Object Right, are larger in the ASC group (Fig 3).

Table 2 presents the descriptive statistics and results of the Mann–Whitney U test for between-group comparisons of transition propensities for each AOI pair during the SSRs trials, limited to AOI pairs showing statistically significant differences. To account for multiple comparisons, we additionally applied the False Discovery Rate (FDR) correction, and the adjusted p-values are reported. Importantly, the overall pattern of results remained unchanged: in the *Witches* trial, all significant differences were confirmed, and in the *Sheriff* trial, only one comparison slightly exceeded the corrected significance threshold, while the observed trend

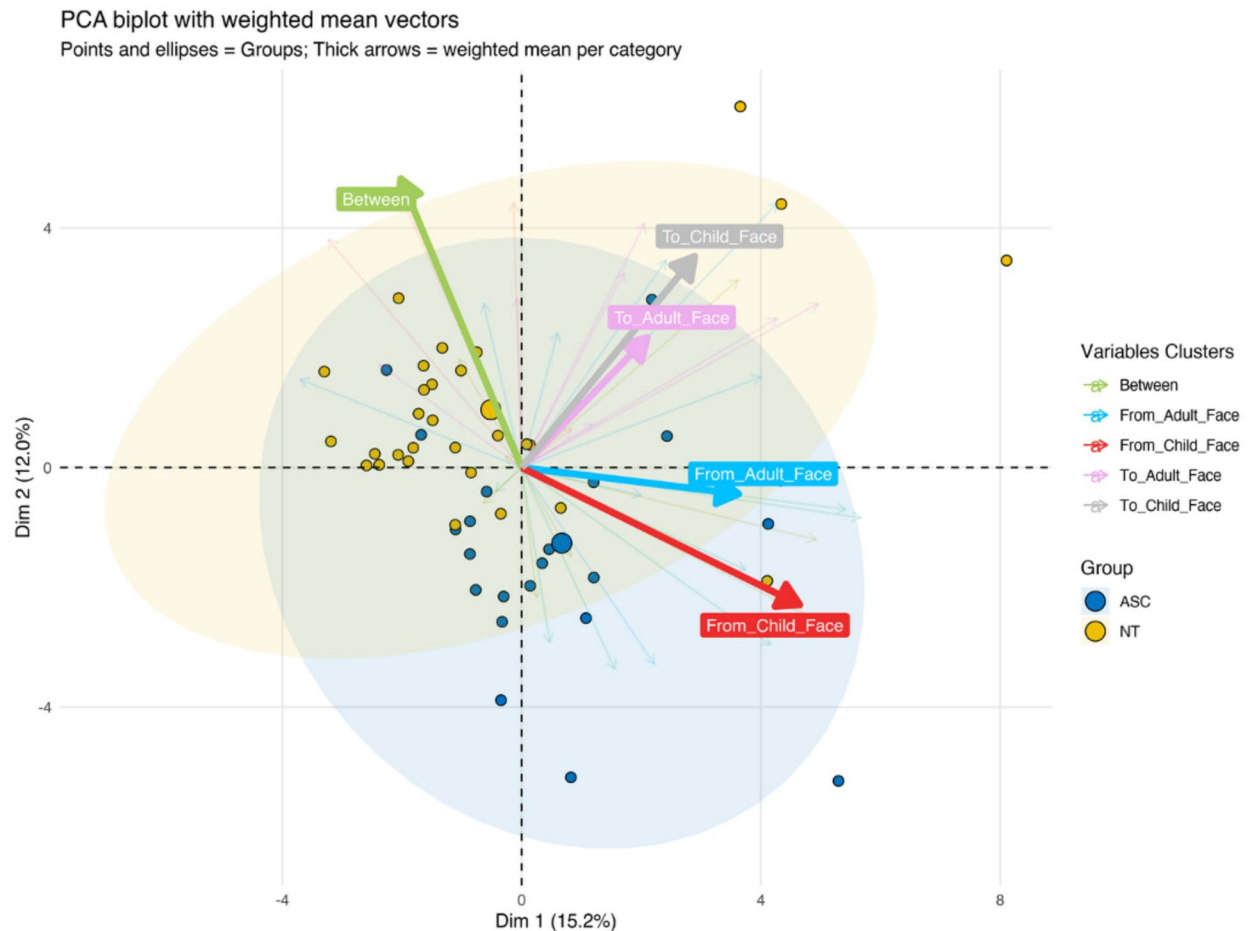


Fig. 2. Scatter plot showing individuals based on the first and second principal components, with concentration ellipses around each group (ASC and NT). Overlaid, a correlation plot displaying the relationship between the first and second principal components, with cluster-averaged loading vectors of transition propensity features, color-coded based on the categorization. Stimulus videos: *Sheriff* and *Witches*.

remained consistent. The full set of comparisons, including non-significant results, is provided in Supplementary Tables S13 (*Sheriff*) and S14 (*Witches*).

During musical activities involving instruments, NT children exhibited gaze patterns characterized by dynamic reorientation toward socially relevant activity areas. Specifically, they frequently redirected their attention from the face of one social partner to the activity area of the other partner, consistent with a gaze triangulation strategy (e.g., from the Adult Face to the Child's Activity AOI in the *Drums* trial). Additionally, NT children demonstrated increased transitions from Distractor Objects (Left or Right) back to the Activity areas, indicating a flexible allocation of attention toward the ongoing shared task. These patterns are visually represented in the chord diagrams by the prominent green arrow from Child Face to Adult Activity, reflecting a higher frequency of cross-partner gaze shifts in the NT group. Additionally, the larger dark gray arrow from Object Right to Adult Face and the light gray arrow from Object Left to Child Face further confirm NT children's greater propensity to reorient their gaze from peripheral distractor objects back to faces (Fig 5).

In contrast, ASC children showed significantly higher gaze transition propensities between the Activity-related AOIs (Adult Activity and Child Activity) compared to NT children, as well as from Activity-related AOIs to distractor objects (specifically, from the Adult Activity AOI to the Object on the Right). This behavior is evident in the chord diagrams, where the larger red arrow from Adult Activity to Child Activity and the yellow arrow from Child Activity to Adult Activity in the *Drums* trial indicate stronger transitions between activity areas (Fig 5). Additionally, ASC children demonstrated a higher propensity to shift gaze from activity-related AOIs to distractor objects, as shown by the more prominent red arrow from Adult Activity to Object Right and the yellow arrow from Child Activity to Object Left in this group. Furthermore, ASC children tend to shift gaze from the face to the Activity-related AOI within the same individual, without directing attention to the other social partner in the scene, thus showing less of the gaze triangulation pattern observed in NT children. These patterns are evident in the PCA space with the average cluster vectors (Fig 4), while the complete feature-level loadings are reported in the Supplementary Materials (Figure S2). Consistently, the chord diagrams (Fig 5) corroborate this behavior, with the green arrow from Child Face to Child Activity and the blue arrow from Adult Face to Adult Activity are significantly larger in the ASC group.

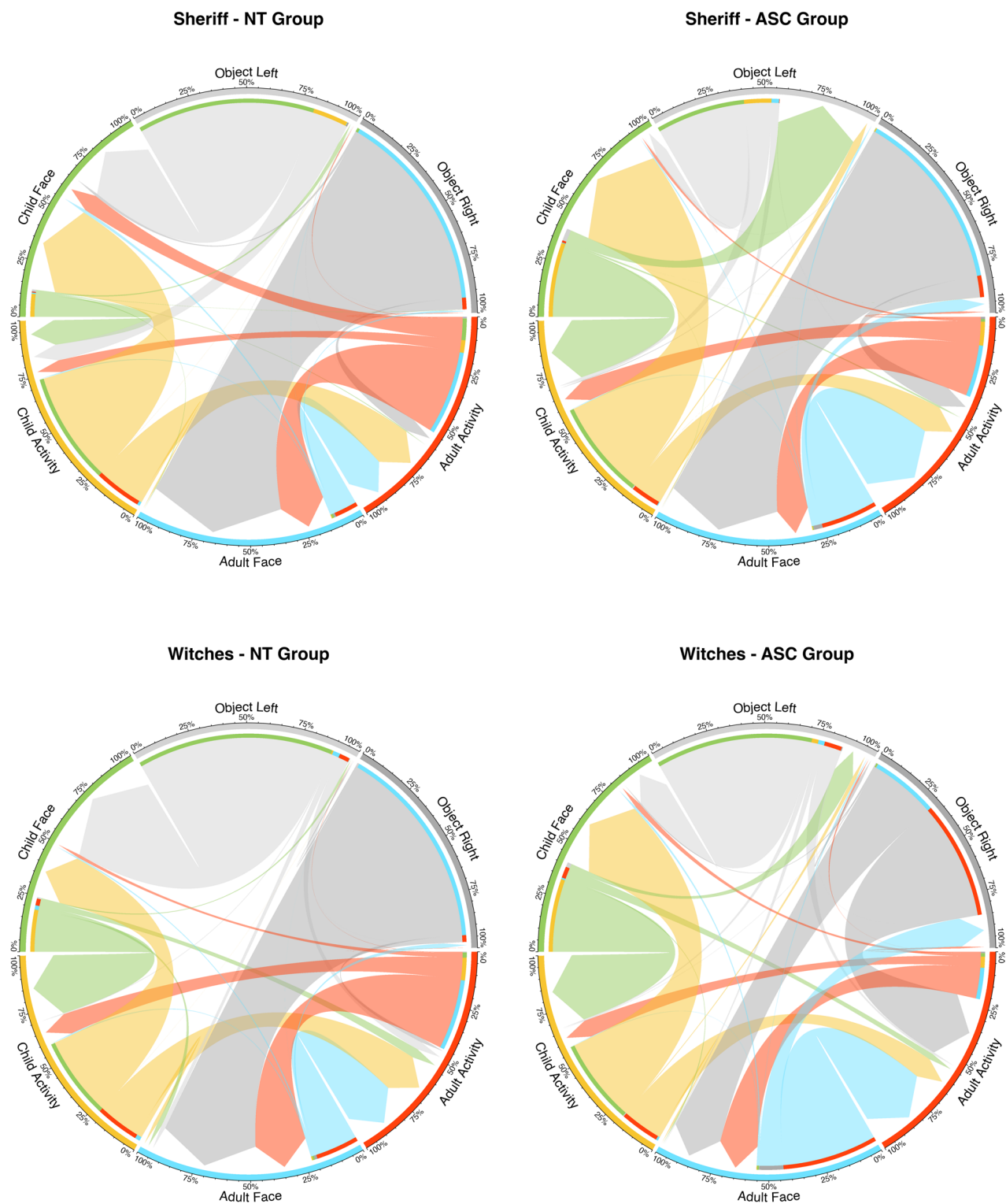


Fig. 3. Chord diagrams representing gaze transition propensities between areas of interest (AOIs) during the *Sheriff* (top) and the *Witches* (bottom) trials. The diagram on the left illustrates transition patterns for the NT group, while the diagram on the right shows those for the ASC group. AOIs are color-coded as follows: Child Face (green), Child Activity (yellow), Adult Face (blue), Adult Activity (red), Object Right (dark gray), and Object Left (light gray). The color of each arrow corresponds to the source AOI, and the arrowhead indicates the target AOI. Arrow thickness indicates the strength of the transitions between AOIs.

Trial	Transition	Group	Mean	Median	SD	Minimum	Maximum	Mann–Whitney U			
								Statistics	p	FDR p	Effect size
Sheriff	Adultactivity—Childface	ASC	3.47e-4	0.00000	4.89e-4	0.00	0.00125	172	0.001	0.002	0.50145
		NT	0.00263	0.00114	0.00446	0.00	0.02000				
	Adultface—Adultactivity	ASC	0.01262	0.01262	0.01185	0.00	0.04343	130	< 0.001	0.008	0.62319
		NT	0.00499	7.12e-4	0.01549	0.00	0.08408				
	Adultface—Childface	ASC	1.60e-4	0.00000	2.73e-4	0.00	9.03e-4	150	< 0.001	0.005	0.56667
		NT	4.83e-4	3.69e-4	3.61e-4	0.00	0.00116				
	Childface—Adultactivity	ASC	2.74e-4	2.27e-4	2.70e-4	0.00	0.00104	227	0.032	0.057	0.34348
		NT	1.29e-4	9.44e-5	1.42e-4	0.00	5.11e-4				
	Childface—Adultface	ASC	7.71e-5	0.00000	1.44e-4	0.00	4.73e-4	165	< 0.001	0.002	0.52174
		NT	1.82e-4	1.70e-4	1.20e-4	0.00	4.99e-4				
	Childface—Childactivity	ASC	0.00964	0.00800	0.01027	6.99e-4	0.04773	127	< 0.001	0.002	0.63188
		NT	0.00311	3.76e-4	0.00583	0.00	0.02607				
	Childface—Objectleft	ASC	0.00148	1.48e-4	0.00493	0.00	0.02387	223	0.024	0.048	0.35362
		NT	2.44e-4	3.77e-5	6.63e-4	0.00	0.00357				
	Objectleft—Childface	ASC	0.00167	0.00152	0.00166	0.00	0.00599	120	< 0.001	0.004	0.65362
		NT	0.01059	0.01059	0.01303	0.00	0.05988				
	Objectright—Adultface	ASC	0.02640	0.02640	0.02254	7.39e-4	0.11765	114	< 0.001	0.003	0.66957
		NT	0.03100	0.03100	0.01271	7.27e-4	0.08237				
Witches	Adultface—Objectright	ASC	4.39e-5	4.39e-5	7.95e-5	0.00	3.95e-4	119	< 0.001	0.016	0.65507
		NT	3.94e-4	3.94e-4	4.01e-4	0.00	0.00236				
	Childface—Adultface	ASC	0.00256	3.29e-4	0.00828	0.00	0.04000	156	< 0.001	0.007	0.54783
		NT	1.36e-4	0.00000	3.55e-4	0.00	0.00178				
	Childface—Objectleft	ASC	1.37e-4	0.00000	1.99e-4	0.00	6.13e-4	206	0.009	0.038	0.40290
		NT	5.11e-4	2.85e-4	5.91e-4	0.00	0.00206				
	Objectleft—Adultface	ASC	4.32e-4	3.12e-4	4.62e-4	0.00	0.00195	211	0.014	0.039	0.38841
		NT	1.92e-4	8.28e-5	2.58e-4	0.00	0.00106				
	Objectright—Adultface	ASC	1.16e-4	0.00000	1.94e-4	0.00	7.29e-4	205	0.008	0.038	0.40580
		NT	4.23e-4	4.69e-4	4.47e-4	0.00	0.00168				
	Objectright—Childface	ASC	0.00603	0.00386	0.00966	0.00	0.04429	211	0.014	0.039	0.38841
		NT	0.01012	0.01010	0.01712	5.52e-4	0.09823				
	Objectright—Childface	ASC	1.11e-4	0.00000	2.84e-4	0.00	0.00137	210	< 0.001	0.007	0.39130
		NT	0.00000	0.00000	0.00000	0.00	0.00000				

Table 2. Descriptive statistics of transition propensities across defined aois in ASC and NT children and Mann–Whitney test results comparing the two groups in sensory social routine trials. Only transitions showing statistically significant differences between ASC and NT children based on the original p-values (<0.05) are reported here; the corresponding FDR-adjusted p-values are also provided for reference. The full set of comparisons, including both uncorrected and FDR-adjusted p-values, is available in Supplementary Tables S13 (*Sheriff*) and S14 (*Witches*).

Table 3 presents the descriptive statistics and results of the Mann-Whitney U test for between-group comparisons of transition propensities across the selected AOIs during the musical activities with instrument trials, limited to AOI pairs showing statistically significant differences. To account for multiple comparisons, we additionally applied the False Discovery Rate (FDR) correction, and the adjusted p-values are reported. Although some comparisons slightly exceeded the corrected threshold (such as the Adult Face → Child Activity and Adult Activity → Object Right transitions in the Drums trial, and the Adult Activity ↔ Child Activity and Adult Activity → Object Right transitions in the Xylophone trial) the overall patterns remain consistent, and key significant differences continue to support the described trends. The full set of comparisons, including non-significant results, is provided in Supplementary Tables S15 (*Drums*) and S16 (*Xylophone*).

Only transitions showing statistically significant differences between ASC and NT children based on the original p-values (<0.05) are reported here; the corresponding FDR-adjusted p-values are also provided for reference. The full set of comparisons, including both uncorrected and FDR-adjusted p-values, is available in Supplementary Tables S15 (*Drums*) and S16 (*Xylophone*).

Furthermore, in the context of musical activities, the PCA visualization shows that gaze shifts in the ASC group are more broadly dispersed along the principal dimensions, suggesting increased variability in gaze-shifting behavior and greater heterogeneity in attentional strategies. This is quantitatively supported by the area of the 95% confidence ellipse, which is substantially larger for the ASC group (136.0) compared to the NT group (35.4), indicating higher dispersion in the gaze profiles of autistic children.

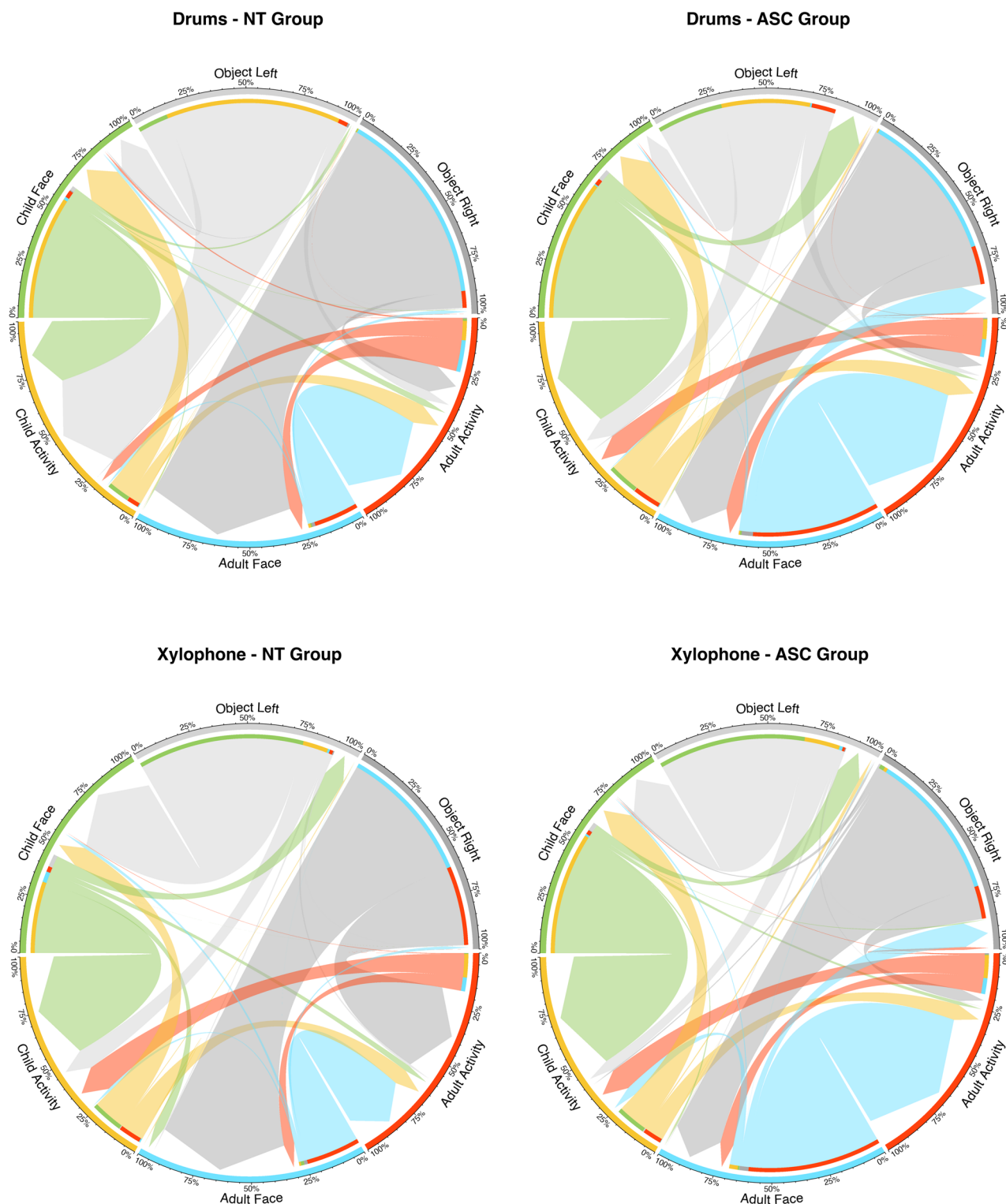


Fig. 5. Chord diagrams representing gaze transition propensities between areas of interest (AOIs) during the *Drums* (top) and the *Xylophone* (bottom) trials. The diagram on the left illustrates transition patterns for the NT group, while the diagram on the right shows those for the ASC group. AOIs are color-coded as follows: Child Face (green), Child Activity (yellow), Adult Face (blue), Adult Activity (red), Object Right (dark gray), and Object Left (light gray). The color of each arrow corresponds to the source AOI, and the arrowhead indicates the target AOI. Arrow thickness indicates the strength of the transitions between AOIs.

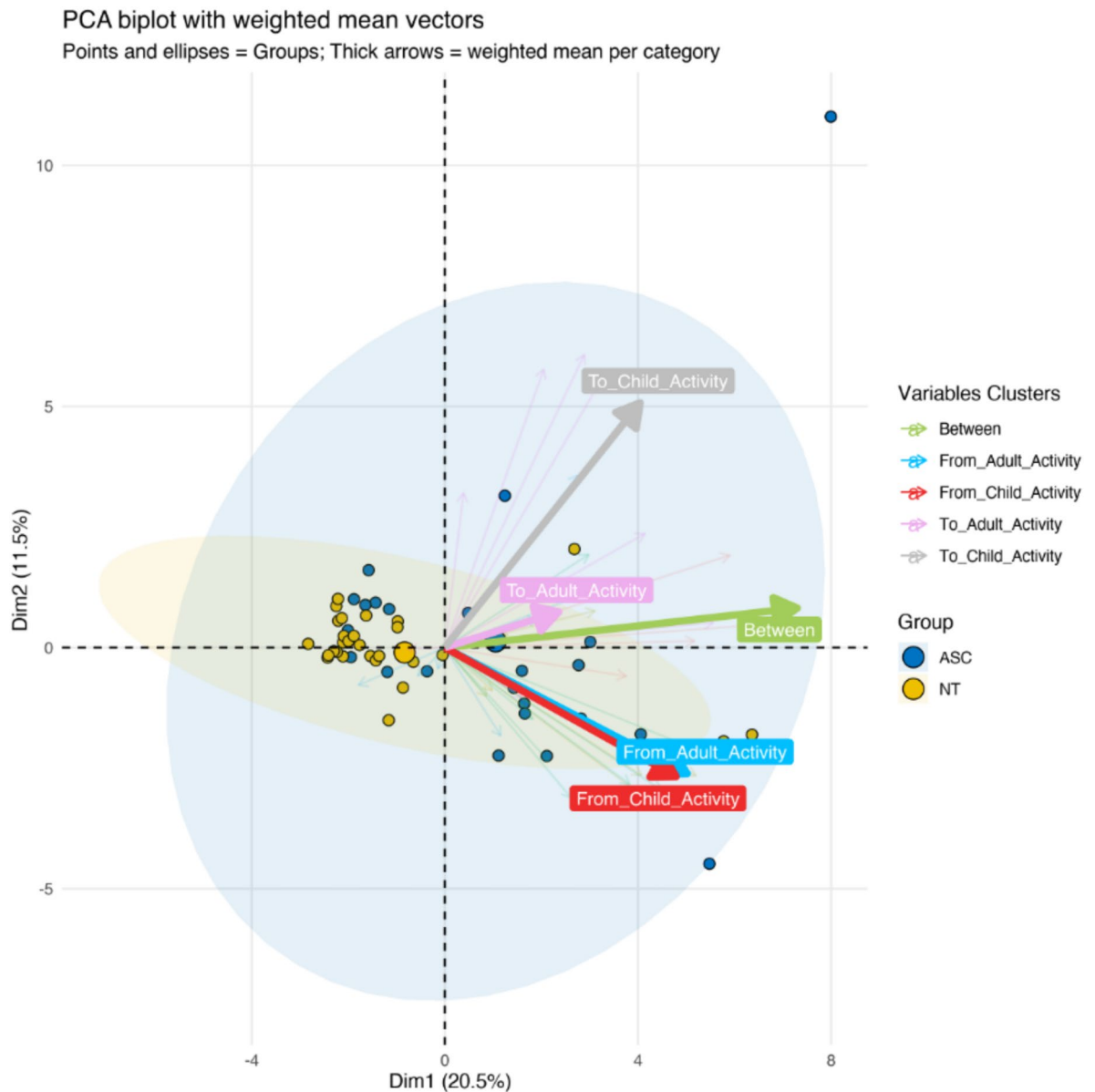


Fig. 4. Scatter plot showing individuals based on the first and second principal components, with concentration ellipses around each group (ASC and NT). Overlaid, a correlation plot displaying the relationship between the first and second principal components, with cluster-averaged loading vectors of transition propensity features, color-coded based on the categorization. Stimulus videos: *Xylophone* and *Drums*.

The MANOVA confirmed a statistically significant group separation. For the SSR trials, the effect of Group was significant on PC2 ($p < 0.001$) and showed a trend on PC1 ($p = 0.065$). For the musical activities, Group significantly explained variance on PC1 ($p = 0.0099$), while no effect was found for PC2 ($p = 0.686$). These results quantitatively support the visual patterns in the PCA plots, reinforcing the robustness of the observed group-level differences in gaze dynamics.

These components reflect distinct behavioral dimensions of gaze dynamics. In the SSR trials, PC2, which showed the strongest group separation, was primarily driven by transitions between the faces of the two partners and by re-engagements from distractor objects toward faces, suggesting that this component captures a social-attention dimension characterized by reciprocal orientation and face-centered coordination. PC1, which showed only a trend-level effect, was largely associated with transitions linking faces and activity areas within the same partner and, to a lesser extent, cross-partner triangulation, reflecting a task-driven integration of social and action-related cues.

In the musical activities, PC1, where group differences were significant, was dominated by cross-partner shifts between the child's and the adult's activity areas and by transitions linking activity zones and faces, capturing a social coordination and triangulation axis. PC2, which did not show significant group separation, loaded more strongly on reorientations from peripheral objects and on within-partner shifts between face and

Trial	Transition	Group	Mean	Median	SD	Minimum	Maximum	Mann–Whitney U			
								Statistics	p	FDR p	Effect size
Drums	Adultactivity—Childactivity	ASC	0.00234	0.00135	0.00239	2.56e-4	0.00864	181	0.004	0.024	0.45727
		NT	0.00109	6.38e-4	0.00121	1.12e-4	0.00524				
	Adultactivity—Objectright	ASC	8.58e-5	0.00000	1.58e-4	0.00000	6.87e-4	238	0.024	0.072	0.28786
		NT	3.47e-5	0.00000	1.08e-4	0.00000	5.05e-4				
	Adultface—Adultactivity	ASC	0.01598	0.00942	0.01806	0.00000	0.07020	133	< 0.001	0.018	0.60120
		NT	0.00566	0.00127	0.01537	3.03e-4	0.08353				
	Adultface—Childactivity	ASC	1.69e-4	0.00000	3.08e-4	0.00000	0.00131	229	0.045	0.115	0.31334
		NT	3.07e-4	2.25e-4	3.25e-4	0.00000	0.00110				
	Childface—Childactivity	ASC	0.01125	0.00831	0.00910	0.00118	0.03012	198	0.012	0.043	0.40630
		NT	0.00674	0.00246	0.00873	3.00e-4	0.03366				
	Objectleft—Childactivity	ASC	0.00219	0.00174	0.00286	0.00000	0.01268	197	0.010	0.043	0.40930
		NT	0.01060	0.01060	0.02043	0.00000	0.11364				
	Objectright—Childactivity	ASC	7.86e-5	0.00000	2.05e-4	0.00000	9.69e-4	183	0.003	0.024	0.45127
		NT	8.27e-5	8.27e-5	1.41e-4	0.00000	7.85e-4				
Xylophone	Adultactivity—Childactivity	ASC	0.00267	0.00195	0.00247	0.00000	0.00988	209	0.021	0.057	0.37331
		NT	0.00165	0.00102	0.00194	4.21e-4	0.00846				
	Adultactivity—Objectright	ASC	1.31e-4	0.00000	4.13e-4	0.00000	0.00197	260	0.023	0.057	0.22039
		NT	1.28e-5	0.00000	6.89e-5	0.00000	3.71e-4				
	Adultface—Adultactivity	ASC	0.01690	0.00653	0.02623	0.00000	0.11905	149	< 0.001	0.005	0.55322
		NT	0.00461	0.00175	0.00836	4.96e-4	0.04150				
	Childactivity—Adultactivity	ASC	0.00202	0.00153	0.00159	2.69e-4	0.00572	212	0.025	0.057	0.36432
		NT	0.00122	6.83e-4	0.00142	0.00000	0.00559				
	Childface—Childactivity	ASC	0.01298	0.00991	0.01355	5.19e-4	0.06497	169	0.002	0.008	0.49325
		NT	0.00592	0.00263	0.00735	0.00000	0.02696				
	Objectright—Adultactivity	ASC	0.00147	0.00144	0.00272	0.00000	0.01355	136	< 0.001	0.005	0.59220
		NT	0.00652	0.00652	0.00849	0.00000	0.04843				
	Objectright—Childactivity	ASC	1.46e-4	7.46e-5	3.66e-4	0.00000	0.00179	160	< 0.001	0.005	0.52174
		NT	0.00000	0.00000	0.00000	0.00000	0.00000				

Table 3. Descriptive statistics of transition propensities across defined aois in ASC and NT children and Mann–Whitney test results comparing the two groups in musical activities trials.

activity, reflecting a profile of peripheral attraction and reduced cross-partner integration. The top 10 loadings contributing to each principal component are provided as bar plots in the Supplementary Materials (Figures S3–S6), highlighting the transitions that most strongly drive PC1 and PC2 in both the SSR and musical activity contexts.

This pattern is consistent with the social and cognitive demands of each interaction type. In the SSR trials, which emphasize dyadic engagement and reciprocal imitation, PC2, capturing face-to-face coordination and re-engagement toward socially salient cues, emerges as the most discriminative component. Conversely, in the musical activity trials, where triadic coordination and task-driven attention are required, PC1, reflecting cross-partner activity integration and face-activity triangulation, plays a more prominent role in explaining group-level differences.

Comparison with dwell time

As a descriptive benchmark, dwell-time analyses showed broad group differences in overall allocation of visual attention across AOIs. Relative to neurotypical (NT) peers, autistic (ASC) children spent less time on faces and more time on non-social elements. Specifically, Adult Face: ASC $11.28\% \pm 10.35$ vs. NT $23.19\% \pm 16.32$; Child Face: ASC $14.35\% \pm 14.57$ vs. NT $20.70\% \pm 20.27$. Conversely, ASC children devoted more time to Adult Activity ($24.63\% \pm 14.16$ vs. $20.98\% \pm 13.88$), Child Activity ($23.74\% \pm 17.36$ vs. $19.50\% \pm 14.48$), Left Object ($4.62\% \pm 9.12$ vs. $1.82\% \pm 2.98$), Right Object ($2.45\% \pm 4.36$ vs. $0.89\% \pm 2.26$), and Background ($18.91\% \pm 14.01$ vs. $12.89\% \pm 10.75$). Details of dwell-time allocation for each individual trial are reported in Figs 6 and 7.

These static differences confirm a general shift away from socially salient regions; however, they do not capture the temporal organization of gaze. By contrast, the CTMC results reveal group-specific sequential patterns (e.g., face→face coordination in SSRs; cross-partner triangulation in instrument trials) that are not explained by dwell time alone. This benchmark therefore clarifies the added value of the Markov-chain approach in characterizing the dynamics of visual exploration.

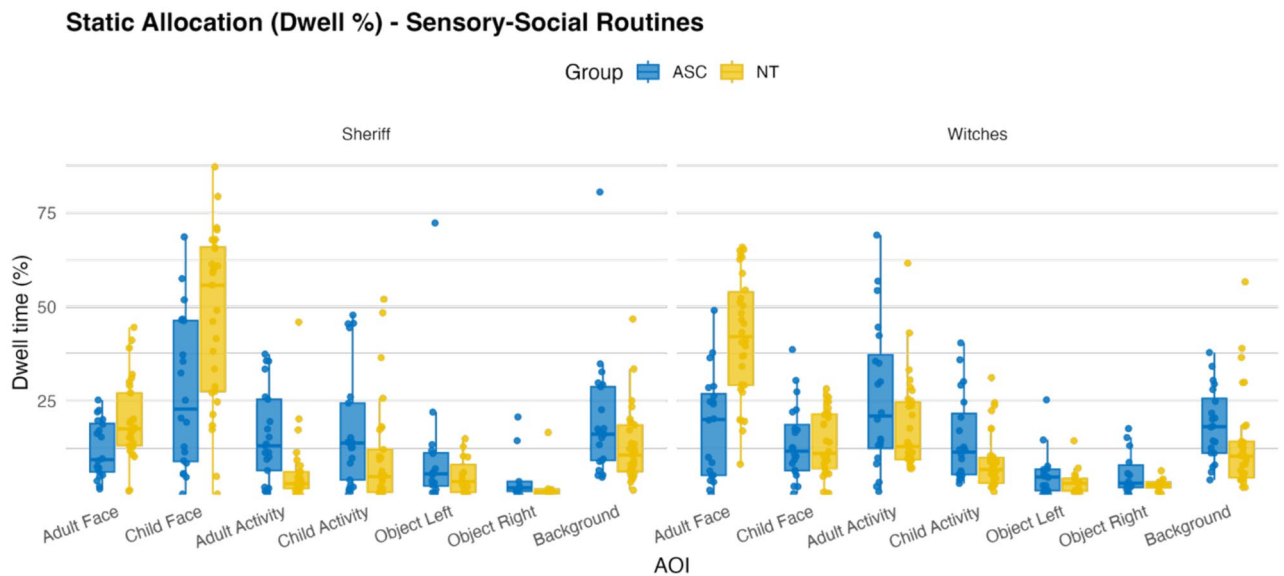


Fig. 6. Static allocation of visual attention (dwell time, %) across Areas of Interest (AOIs) during the Sensory-Social Routine (SSR) trials (*Sheriff* and *Witches*). Boxplots represent the proportion of total looking time directed to each AOI, separately for autistic (ASC, blue) and neurotypical (NT, yellow) children. Boxes show the interquartile range (IQR), horizontal lines indicate the median, whiskers extend to $1.5 \times \text{IQR}$, and dots represent individual participants.

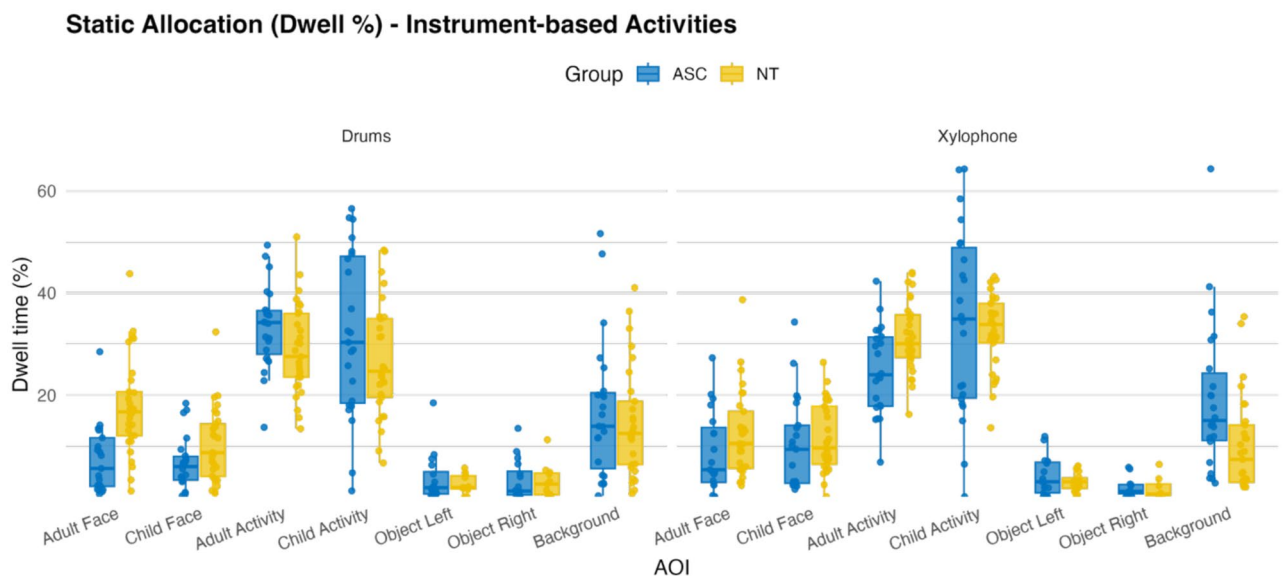


Fig. 7. Static allocation of visual attention (dwell time, %) across Areas of Interest (AOIs) during the instrument-based musical activity trials (*Drums* and *Xylophone*). Boxplots depict the proportion of total looking time allocated to each AOI for autistic (ASC, blue) and neurotypical (NT, yellow) children. Boxes show the interquartile range (IQR), horizontal lines indicate the median, whiskers extend to $1.5 \times \text{IQR}$, and dots represent individual participants.

Discussion

This study adopted an innovative analytical framework combining Continuous-Time Markov Chains (CTMCs) and Principal Component Analysis (PCA) to investigate the temporal structure of gaze shifts between social and non-social elements during the observation of ecologically valid, play scenarios involving dyadic and triadic interactions. By moving beyond traditional static measures of gaze allocation to predefined Areas of Interest (AOIs), this approach captured the sequential dynamics of visual attention in young children with and without autism, while reducing data complexity and facilitating a clearer interpretation of patterns often obscured in aggregate dwell-time analyses.

A major methodological contribution of this study lies in formalizing these behaviors as time-dependent transitions between AOIs. By modeling gaze as a probabilistic process, the CTMC framework enables quantification of attentional shifts and dwell times, offering a high-resolution view of how gaze behaviors unfold dynamically over time and revealing subtle differences in re-engagement tendencies often overlooked by static measures. Compared to traditional discrete-time HMM approaches, which infer temporal dynamics only indirectly, the continuous-time formulation directly parameterizes both transition rates and dwell times³², providing richer and more interpretable descriptors of attentional organization. However, the method entails trade-offs: the “memoryless” assumption simplifies estimation but may underrepresent longer-range attentional dependencies, and the PCA step, while valuable to capture inter-individual variability, adds an interpretive layer less intuitive for immediate clinical translation. For these reasons, this framework should be considered complementary to simpler percentage- or dwell-time-based metrics and hypothesis-driven analyses, with each approach providing synergistic insights into attentional organization.

The analyses revealed context-dependent gaze strategies across groups. In the sensory social routines (SSRs), characterized by dyadic imitation and song-based engagement, NT children showed a robust social orientation, marked by frequent gaze transitions between faces and consistent reorienting toward face-related AOIs following distraction. These shifts likely support synchronization and joint attention, foundational for social development, and reflect a prioritization of socially salient elements during exploration, consistent with prior evidence of an attentional bias toward faces in neurotypical development that supports early social learning and communicative reciprocity^{3,40–47}.

In contrast, ASC children exhibited a divergent attentional strategy, with a greater tendency to disengage from socially salient cues and redirect gaze toward non-social elements, and reduced re-engagement with faces after these shifts. This pattern aligns with evidence of decreased social preference and a prioritization of object-based or peripheral stimuli^{2,16,17,48}, potentially reflecting underlying differences in social motivation^{4,49–51} or adaptive responses to social overstimulation. Moreover, their exploration dynamics mirrored previous findings of a continuous decline in social attention and lack of re-engagement with faces toward the end of the trial^{10,21,22,52,53}.

In the more complex musical instrument trials, characterized by triadic interactions parallel object use, and synchronized turn-taking, NT children showed a coherent attentional strategy, focusing on activity areas and consistently reorienting to task-relevant regions after distraction, consistent with previous research^{54,55}. They also displayed consistent gaze triangulation, shifting attention between a partner’s face and the other partner’s activity, supporting dynamic coordination of attention across socially relevant components in the scene.

Conversely, ASC children engaged in more fragmented visual exploration, with higher transition propensities between activity areas, frequent within-actor gaze shifts, and more distractions from task-relevant zones to peripheral elements. This indicates reduced integration of social engagement cues and a less socially-driven exploration style, consistent with evidence that ASC toddlers are more easily distracted by background objects^{54,55}, and that their attentional patterns are guided more by object features or peripheral salience than by social coordination.

The analysis of inter-individual variability through PCA further highlighted heterogeneity within the ASC group, particularly during musical instrument trials, as reflected by a more diffuse distribution and a larger 95% confidence ellipse. This finding aligns with previous work documenting idiosyncratic gaze behaviors in autism^{56–58} and with evidence linking greater variability in gaze dynamics to reduced scene comprehension⁵⁹ and higher autistic trait expression⁶⁰.

Importantly, the unsupervised PCA approach is not a “black box” but a descriptive tool that summarizes complex transition profiles into a few dimensions that can be meaningfully interpreted. In our data, the primary components reflected contrasts between social re-engagement and peripheral capture, and between coordinated triadic attention and more fragmented, within-actor tracking. These components correlated with simpler indices, such as face-looking percentages and number of re-orientations, underscoring that PCA captures higher-order combinations of familiar gaze behaviors and providing complementary insights beyond those offered by aggregate metrics.

By integrating CTMC and PCA within a naturalistic paradigm, this study advances the methodological toolkit for neurodevelopmental research. The framework not only detects group-level patterns but also quantifies individual variability, providing a promising foundation for the development of dynamic, process-based biomarkers of social attention.

Beyond methodological innovation, the approach holds strong clinical potential: by capturing fine-grained temporal dynamics, it could support early screening of children at risk for autism, longitudinal monitoring of developmental changes or intervention effects, and the personalization of interventions tailored to each child’s attentional profile. Validation in larger, multi-site cohorts will be critical to establish predictive value and clinical utility.

Among the key strengths of this study are the use of ecologically valid, socially rich stimuli mimicking naturalistic interactions and the inclusion of both dyadic (SSRs) and triadic (musical) contexts, allowing a nuanced comparison of gaze dynamics under varying social demands. Additionally, the probabilistic modeling of gaze transitions provided a comprehensive view of attentional organization, and the robustness of the main group-level patterns, even after correction for multiple comparisons, reinforces the reliability of these findings. Future studies should aim to replicate these results in larger, demographically balanced samples and narrower age bands to improve generalizability and capture developmental changes.

Nevertheless, the present study has several limitations. The modest sample size and group heterogeneity may limit generalizability, as may the gender imbalance between groups and the broad age range of participants. Moreover, the AOIs, while capturing the main social and non-social elements, did not encode stimulus-level interaction events, such as moments of joint attention, which could be included in future extensions to link gaze dynamics more directly with interaction events. Another important aspect concerns the sensitivity of this

approach to data quality: eye-tracking is prone to signal noise, track loss, or brief off-screen episodes, which can distort transition counts or dwell estimates. Although preprocessing steps (e.g., merging adjacent fixations, excluding low-confidence samples, encoding track-loss as a dedicated state) helped mitigate these issues, estimates may still be less stable in noisier datasets. Future applications would benefit from formal robustness checks, such as varying fixation thresholds or AOI boundaries, or hierarchical modeling to stabilize parameter estimation. Finally, by focusing on transition probabilities rather than exact temporal sequences, some information on the order of gaze events is not preserved; integrating complementary methods such as sequence or event-based analyses could enrich future investigations.

Importantly, the proposed framework opens promising avenues for longitudinal investigations of gaze behavior as a developmental marker, supporting both early detection and individualized intervention strategies in autism.

Conclusion

This study combined continuous-time Markov chains (CTMCs) and principal component analysis (PCA) to characterize the temporal architecture of gaze transitions between social and non-social elements in young autistic and non-autistic children during naturalistic interactive play. By adopting a dynamic modeling framework, the study moved beyond traditional static gaze metrics, capturing the fine-grained structure and variability of visual attention in ecologically valid scenarios. The experimental paradigm included both sensory social routines (SSRs) involving song-based dyadic interactions and object-based musical activities with instruments requiring triadic coordination.

Findings revealed distinct group-level gaze strategies. NT children consistently demonstrated social attentional preferences, characterized by frequent gaze alternations between faces and reorientation from non-social stimuli back to social partners, particularly during SSRs. These patterns reflect a typical developmental trajectory that prioritizes social engagement and supports the emergence of joint attention and reciprocal interaction. In contrast, autistic children more often disengaged from faces, directing attention toward peripheral or non-social elements such as distractor objects and activity zones. They also exhibited higher transition frequencies between non-social AOIs and reduced gaze triangulation, suggesting differences in attentional allocation and coordination in social contexts.

These findings contribute to a growing body of research highlighting the variability and complexity of visual attention in autism. The use of CTMCs allowed for a detailed mapping of moment-to-moment gaze transitions, while by projecting individual gaze transition profiles into the PCA space, we were able to capture differences in their distribution, providing insight into inter-individual variability within and across groups. This dual-method approach enabled the identification of subtle, yet meaningful, differences in how autistic and non-autistic children visually explore socially embedded contexts.

Importantly, the results underscore the relevance of considering both the structure and dynamics of gaze behavior when studying neurodevelopmental conditions. The divergent patterns observed in autistic children may reflect reduced social motivation, differences in scanning strategies, or adaptive responses to heightened sensory input. Gaining a deeper understanding of these processes is crucial for informing the design of supportive environments and interventions that honor neurodivergent profiles and foster meaningful participation in naturalistic settings.

Further research is needed to clarify the mechanisms underlying these gaze behaviors, in order to guide the development of targeted approaches that strengthen opportunities for joint attention and reciprocal engagement. Such approaches should be attuned to the distinctive needs and preferences of neurodivergent autistic children within real-world social contexts.

Looking forward, this dynamic, process-based approach holds promise for early identification, longitudinal monitoring, and the personalization of interventions for autistic children, supporting a more nuanced and individualized understanding of social attention trajectories in neurodevelopment.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Received: 9 July 2025; Accepted: 6 October 2025

Published online: 12 November 2025

References

1. Elsabbagh, M. et al. Social and attention factors during infancy and the later emergence of autism characteristics. *Prog. Brain. Res.* **189**, 195–207 (2011).
2. Falck-Ytter, T., Kleberg, J. L., Portugal, A. M. & Thorup, E. Social Attention: Developmental Foundations and Relevance for Autism Spectrum Disorder. *Biol. Psychiatry* **94**, 8–17 (2023).
3. Klin, A., Shultz, S. & Jones, W. Social visual engagement in infants and toddlers with autism: early developmental transitions and a model of pathogenesis. *Neurosci. Biobehav. Rev.* **50**, 189–203 (2015).
4. Moriuchi, J. M., Klin, A. & Jones, W. Mechanisms of diminished attention to eyes in autism. *Am. J. Psychiatry* **174**, 26–35 (2017).
5. Riby, D. M. & Hancock, P. J. B. Viewing it differently: social scene perception in williams syndrome and autism. *Neuropsychologia* **46**, 2855–2860 (2008).
6. Sasson, N. J. The development of face processing in autism. *J. Autism Dev Disord* **36**, 381–394 (2006).
7. Sasson, N. J., Pinkham, A. E., Weittenhiller, L. P., Faso, D. J. & Simpson, C. Context effects on facial affect recognition in schizophrenia and autism: behavioral and eye-tracking evidence. *Schizophr. Bull.* **42**, 675–683 (2016).
8. Wilson, C. E., Palermo, R. & Brock, J. Visual scan paths and recognition of facial identity in autism spectrum disorder and typical development. *PLoS ONE* **7**, e37681 (2012).

9. Frazier, T. W. et al. A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *J. Am. Acad. Child. Adolesc. Psychiatry* **56**, 546–555 (2017).
10. Kojovic, N. et al. Unraveling the developmental dynamic of visual exploration of social interactions in autism. *Elife* **13**, e85623 (2024).
11. Moore, A. et al. The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking. *Mol. Autism* **9**, 19 (2018).
12. Pierce, K., Conant, D., Hazin, R., Stoner, R. & Desmond, J. Preference for geometric patterns early in life as a risk factor for autism. *Arch. Gen. Psychiatry* **68**, 101–109 (2011).
13. Pierce, L. J., Chen, J.-K., Delcenserie, A., Genesee, F. & Klein, D. Past experience shapes ongoing neural patterns for language. *Nat. Commun.* **6**, 10073 (2015).
14. Shi, L. et al. Different visual preference patterns in response to simple and complex dynamic social stimuli in preschool-aged children with autism spectrum disorders. *PLoS ONE* **10**, e0122280 (2015).
15. Chevallier, C. et al. Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Res* **8**, 620–628 (2015).
16. Hedger, N., Dubey, I. & Chakrabarti, B. Social orienting and social seeking behaviors in ASD. A meta analytic investigation. *Neurosci. Biobehav. Rev* **119**, 376–395 (2020).
17. Hou, W., Jiang, Y., Yang, Y., Zhu, L. & Li, J. Evaluating the validity of eye-tracking tasks and stimuli in detecting high-risk infants later diagnosed with autism: A meta-analysis. *Clin. Psychol. Rev.* **112**, 102466 (2024).
18. Saitovitch, A. et al. Studying gaze abnormalities in autism: Which type of stimulus to use? **2013**, (2013).
19. Verneti, A. et al. Face-to-face live eye-tracking in toddlers with autism: feasibility and impact of familiarity and face covering. *Autism. Res.* **17**, 1381–1390 (2024).
20. Falck-Ytter, T. Gaze performance during face-to-face communication: a live eye tracking study of typical children and children with autism. *Res. Autism. Spectr. Disord.* **17**, 78–85 (2015).
21. Hedger, N. & Chakrabarti, B. Autistic differences in the temporal dynamics of social attention. *Autism* **25**, 1615–1626 (2021).
22. Del Bianco, T. et al. Temporal profiles of social attention are different across development in autistic and neurotypical people. *Biol. Psychiatry: Cognitive Neurosci. Neuroim* **6**, 813–824 (2021).
23. Chuk, T., Chan, A. B. & Hsiao, J. H. Understanding eye movements in face recognition using hidden Markov models. *J. Vis.* **14**, 8 (2014).
24. Coutrot, A., Hsiao, J. H. & Chan, A. B. Scanpath modeling and classification with hidden Markov models. *Behav. Res.* **50**, 362–379 (2018).
25. Hsiao, J. H., Lan, H., Zheng, Y. & Chan, A. B. Eye movement analysis with hidden Markov models (EMHMM) with co-clustering. *Behav. Res. Methods* **53**, 2473–2486 (2021).
26. Kim, J., Singh, S., Thiessen, E. D. & Fisher, A. V. A hidden Markov model for analyzing eye-tracking of moving objects. *Behav. Res.* **52**, 1225–1243 (2020).
27. Ulutas, B. H., Özkan, N. F. & Michalski, R. Application of hidden markov models to eye tracking data analysis of visual quality inspection operations. *Cent. Eur. J. Oper. Res.* **28**, 761–777 (2020).
28. Ma, L. The Application of Hidden Markov Model in the Eye Movement Data. in *Proceedings of The fourth International Conference on Information Science and Cloud Computing — PoS(ISCSC2015)* 042 (Sissa Medialab, Guangzhou, China, 2016). <https://doi.org/10.22323/1.264.0042>.
29. Chen, Z. et al. Process mining IPTV customer eye gaze movement using discrete-time markov chains. *Algorithms* **16**, 82 (2023).
30. Boccignone, G. (2019) Advanced statistical methods for eye movement analysis and modeling: a gentle introduction. in 309–405 https://doi.org/10.1007/978-3-030-20085-5_9.
31. Zhu, Y., Yan, Y., Komogortsev, O. (2020) Hierarchical HMM for Eye Movement Classification. Preprint at <https://doi.org/10.48550/arXiv.2008.07961>
32. Henning, L. X. & Camerer, C. *Estimating hidden markov models (HMMs) of the cognitive process in strategic thinking using eye-tracking 2* (Front. Behav. 2023).
33. Yu, L., Wang, Z., Fan, Y., Ban, L. & Mottron, L. Autistic preschoolers display reduced attention orientation for competition but intact facilitation from a parallel competitor: Eye-tracking and behavioral data. *Autism* **28**, 1551–1564 (2024).
34. Wall, C. A., Shic, F., Varanasi, S. & Roberts, J. E. Distinct social attention profiles in preschoolers with autism contrasted to fragile x syndrome. *Autism Res* **16**, 340–354 (2023).
35. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5TM*, 5th Ed. xlv, 947 (American Psychiatric Publishing, Inc., Arlington, VA, US, 2013). <https://doi.org/10.1176/appi.books.9780890425596>.
36. McCrimmon, A., Rostad, K. Test Review: Lord C, Luyster R J, Gotham K, Guthrie, W. (2012). 'Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module.' Torrance, CA: Western Psychological Services, 2012. Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. 'Autism Diagnostic Observation Schedule, Second Edition.' Torrance, CA: Western Psychological Services, 2012. *J. Psychoeducational Assessment* **32**, 88–92
37. Chawarska, K., Ye, S., Shic, F. & Chen, L. Multilevel differences in spontaneous social attention in toddlers with autism spectrum disorder. *Child. Dev.* **87**, 543–557 (2016).
38. Ching, W.-K., Huang, X., Ng, M. K. & Siu, T. K. *Markov chains: models, algorithms and applications* (springer publishing company, 2013).
39. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A Math Phys Eng Sci* **374**, 20150202 (2016).
40. Bertenthal, B. & Boyer, T. The development of social attention in human infants. *Many Faces Soc. Attention: Behav. Neural Meas.* https://doi.org/10.1007/978-3-319-21368-2_2 (2015).
41. Farroni, T., Csibra, G., Simion, F. & Johnson, M. H. Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9602–9605 (2002).
42. Farroni, T. et al. Newborns' preference for face-relevant stimuli: effects of contrast polarity. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17245–17250 (2005).
43. Frank, M. C., Vul, E. & Johnson, S. P. Development of infants' attention to faces during the first year. *Cognition* **110**, 160–170 (2009).
44. Frank, M. C., Vul, E. & Saxe, R. Measuring the development of social attention using free-viewing. *Infancy* **17**, 355–375 (2012).
45. Johnson, M. H. & Karmiloff-Smith, A. Neuroscience Perspectives on Infant Development. in *Theories of Infant Development* 121–141 (John Wiley & Sons, Ltd, 2004). <https://doi.org/10.1002/9780470752180.ch5>.
46. Simion, F., Macchi Cassia, V., Turati, C. & Valenza, E. The origins of face perception: specific versus non-specific mechanisms. *Infant Child Dev.* **10**, 59–65 (2001).
47. Vivanti, G. & Rogers, S. J. Autism and the mirror neuron system: insights from learning and teaching. *Philosophical. Trans. Royal Soc. B: Biol. Sci.* **369**, 20130184 (2014).
48. Chita-Tegmark, M. Attention allocation in ASD: a review and meta-analysis of eye-tracking studies. *Rev. J. Autism Dev. Disord.* **3**, 209–223 (2016).
49. Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S. & Schultz, R. T. The social motivation theory of autism. *Trends Cogn. Sci.* **16**, 231–239 (2012).

50. Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J. & Brown, E. Children with autism fail to orient to naturally occurring social stimuli. *J. Autism. Dev. Disord.* **28**, 479–485 (1998).
51. Dawson, G., Webb, S. J. & McPartland, J. Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *HDVN* **27**, 403–424 (2005).
52. Kleberg, J. L. et al. Autistic traits and symptoms of social anxiety are differentially related to attention to others' eyes in social anxiety disorder. *J Autism Dev. Disord.* **47**, 3814–3821 (2017).
53. Ni, W., Lu, H., Wang, Q., Song, C. & Yi, L. Vigilance or avoidance: How do autistic traits and social anxiety modulate attention to the eyes?. *Front. Neurosci.* **16**, 1081769 (2023).
54. Harrop, C. et al. Visual attention to faces in children with autism spectrum disorder: are there sex differences?. *Mol. Autism* **10**, 28 (2019).
55. Shic, F., Bradshaw, J., Klin, A., Scassellati, B. & Chawarska, K. Limited activity monitoring in toddlers with autism spectrum disorder. *Brain Res.* **1380**, 246–254 (2011).
56. Avni, I. et al. Children with autism observe social interactions in an idiosyncratic manner. *Autism Res.* **13**, 935–946 (2020).
57. Nakano, T. et al. (2010) Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. *Proceedings of the Royal Society B: Biological Sciences* **277**, 2935–2943
58. Wang, K. K. et al. An update on diagnostic and prognostic biomarkers for traumatic brain injury. *Expert Rev. Mol. Diagn.* **18**, 165–180 (2018).
59. Byrge, L., Dubois, J., Tyszka, J. M., Adolphs, R. & Kennedy, D. P. Idiosyncratic brain activation patterns are associated with poor social comprehension in autism. *J. Neurosci.* **35**, 5837–5850 (2015).
60. Bolton, T. A. W., Freitas, L. G. A., Jochaut, D., Giraud, A.-L. & Van De Ville, D. Neural responses in autism during movie watching: Inter-individual response variability co-varies with symptomatology. *Neuroimage* **216**, 116571 (2020).

Acknowledgements

Simona Campisi is a PhD student enrolled in the National PhD in Artificial Intelligence, XXXVIII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma. The work of Prof. Andrea De Gaetano was supported by the Distinguished Professor Excellence Program of Óbuda University, Budapest Hungary.

Author contributions

R.B., F.I.F., and S.C. wrote the original draft. R.B., F.I.F., L.S., E.L., C.C., S.A., A.C., the NEST Team, and M.M. contributed to data curation and investigation. R.B., L.S., F.I.F., S.C., A.B., and G.T. performed formal analysis and/or visualization. R.B., F.I.F., L.S., B.C., L.R., A.D.G., A.B., and G.T. contributed to methodology and software development. B.C., L.R., A.D.G., and G.T. were responsible for conceptualization and supervision. L.R. and G.T. handled project administration and resource management. B.C., G.P., and G.T. acquired funding. L.R. and G.T. wrote and edited the reviewed version of the manuscript. All authors reviewed and approved the final manuscript.

Funding

Fundings: DREAM project (PRIN—MUR, DSB.AD008.649); READS project (MIMIT, Prog. n. F/180026/01–04/X43); INTER PARES project (EU POC METRO 2014/2020, Azione I.3.1.—Codice Progetto ME I.3.1.b.). Bhismadev Chakrabarti acknowledges funding support from the Medical Research Council UK (Grant ref: MR/S036423/1); and the European Research Council (Grant ref: 865568).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-23366-4>.

Correspondence and requests for materials should be addressed to L.R. or G.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

The NEST Team

M. Boncoddò, I. Crimi, F. Di Bella, G. Lazzaro, S. Martines, A. Orsina & M. Rosano