Synerg-eye-zing: Decoding Nonlinear Gaze Dynamics Underlying Successful Collaborations in Co-located Teams

G. S. Rajshekar Reddy
rajshekar.gudasubhash@colorado.edu
University of Colorado Boulder
Boulder, CO, United States

Lucca Eloy
lucca.eloy@colorado.edu
University of Colorado Boulder
Boulder, CO, United States

Rachel Dickler
rachel.dickler@colorado.edu
University of Colorado Boulder
Boulder, CO, United States

Jason G. Reitman
jason.reitman@colorado.edu
University of Colorado Boulder
Boulder, CO, United States

Samuel L. Pugh
samuel.pugh@colorado.edu
University of Colorado Boulder
Boulder, CO, United States

Peter W. Foltz
peter.foltz@colorado.edu
University of Colorado Boulder
Boulder, CO, United States

Jamie C. Gorman
jamie.gorman@asu.edu
Arizona State University
Mesa, AZ, United States

Julie L. Harrison
julieharrison@gatech.edu
Georgia Institute of Technology
Atlanta, GA, United States

Leanne Hirshfield
leanne.hirshfield@colorado.edu
University of Colorado Boulder
Boulder, CO, United States

ABSTRACT

Joint Visual Attention (JVA) has long been considered a critical component of successful collaborations, enabling coordination and construction of a shared knowledge space. However, recent studies challenge the notion that JVA alone ensures effective collaboration. To gain deeper insights into JVA's influence, we examine nonlinear gaze coupling and gaze regularity in the collaborators' visual attention. Specifically, we analyze gaze data from 19 dyadic and triadic teams engaged in a co-located programming task using Recurrence Quantification Analysis (RQA). Our results emphasize the significance of team-level gaze regularity for improving task performance - highlighting the importance of maintaining stable or sustained episodes of joint or individual attention, than disjointed patterns. Additionally, through regression analyses, we examine the predictive capacity of recurrence metrics for subjective traits such as social cohesion and social loafing, revealing unique interpersonal and team dynamics behind productive collaborations. We elaborate on our findings via qualitative anecdotes and discuss their implications in shaping real-time interventions for optimizing collaborative success.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; *Computer supported cooperative work*.

KEYWORDS

Recurrence quantification analysis; Eye tracking; Group programming; Collaborative problem solving

ICMI '23, October 09–13, 2023, Paris, France

ACM ISBN 979-8-4007-0055-2/23/10.
https://doi.org/10.1145/3577190.3614104

ACM Reference Format:
G. S. Rajshekar Reddy, Lucca Eloy, Rachel Dickler, Jason G. Reitman, Samuel L. Pugh, Peter W. Foltz, Jamie C. Gorman, Julie L. Harrison, and Leanne Hirshfield. 2023. Synerg-eye-zing: Decoding Nonlinear Gaze Dynamics Underlying Successful Collaborations in Co-located Teams. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '23), October 09–13, 2023, Paris, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3577190.3614104

1 INTRODUCTION

In today's world, effective collaboration is crucial to tackle complex challenges and harness collective intelligence. Collaboration is also increasingly emphasized as a predictor of success in higher education and the workforce [15, 35]. Collaborative Problem Solving (CPS), in particular, is when two or more people coordinate to find a solution to a problem [45]. Nevertheless, assessing the skills involved in collaboration and CPS is difficult due to their complexity, involving cognitive and social processes among multiple people [21]. To accurately quantify and understand CPS, it is crucial to employ dynamic multimodal analytics capable of extracting valuable insights into the underlying processes driving successful teamwork.

Furthermore, it is imperative to consider not only objective outcomes such as task performance, but also the subjective qualities crucial for fruitful collaborations. These qualities encompass aspects such as interaction quality, social cohesion, and trust [52]. Social loafing is another factor that can affect group performance [67], a phenomenon where individuals exert less effort in a group setting than when working alone, leading to suboptimal team performance [3, 30]. Hence, further research is needed on the computational assessment of these traits and their potential interventions.

Gaze-based measurement of CPS. Fortunately, an extensive body of research has investigated the feasibility of using data from various modalities (e.g., speech, gesture, eye gaze, physiology; see review [43]) to analyze CPS. Of these modalities, eye gaze has emerged as a promising data stream that is integral in developing the social cognition favorable for collaboration. It provides fine-grained information about where collaborators are looking (e.g., areas of interest (AOI)), as well as indices of social visual attention such as mutual gaze (when teammates look at each other). Consequently, eye-tracking has been utilized in a number of studies on CPS, and eye gaze has been shown to predict important outcomes of CPS such as learning gains [39, 47] and task performance [8, 51], as well as CPS processes such as shared knowledge construction and negotiation [1, 59]. Visual attention was also the most impactful modality (among others) shared between the collaborators, and had the most influence on team synchrony [37].

Moreover, eye gaze is important in CPS as it allows an individual to direct another's visual attention and consequently communicate their intent, thoughts, and desires [18]. Upon perceiving the gaze direction of another individual, one can engage in Joint Visual Attention (JVA) which occurs when both their visual attention aligns on a common object [49]. High levels of JVA have long been discussed as a crucial component to enabling successful collaborations [4, 7, 14, 42], improving the quality of interactions [27, 28, 38], and enhancing collaborative strength [27, 40]. Nonetheless, recent studies suggest it is insufficient for optimal team performance [47, 56, 57]. Since collaborative strategies constantly evolve based on changing circumstances and team members' strengths, the success of any team cannot be attributed to a single trait. We add to the literature by examining the specific role of JVA in promoting collaborative success. We also investigate the relationship between dynamic gaze patterns and both objective task performance and subjective perceptions of collaboration-related qualities.

Nonlinear Dynamics of Collaboration. Researchers have explored various analytical approaches to quantify gaze patterns from eye-tracking signals in collaborative tasks. One such technique is the gaze overlap percentage, which quantifies the time spent by the collaborators focusing on the same area [41]. Other works have examined leader-follower gaze patterns [23, 41, 44], such as the time delay between a target being referenced by one teammate and another teammate fixating on it.

However, collaborating parties form highly intricate systems due to the dynamic interplay among multiple individuals, each with their own unique characteristics, behaviors, and perspectives, leading to nonlinear shifts in their interactions. To comprehensively capture these nonlinear patterns and structures, they should be studied by employing nonlinear tools such as Recurrence Quantification Analysis (RQA). RQA quantifies repeating patterns of a dynamical system by comparing a time series trajectory to itself, exposing the underlying temporal dynamics [63]. It allows for studying features that would otherwise be hidden in an atemporal analysis that aggregates behavioral data over a time span [10].

Cross-RQA (CRQA) [63] is an extension of RQA that measures *similarity* between two different trajectories. It can reveal gaze coupling and quantify JVA between two time-locked gaze signals and also identify revisitation states where an individual's gaze revisits areas already visited by their partner. Multidimensional RQA (MdRQA) [61], another extension of RQA, measures the *regularity* of multiple time series within a single complex system, thereby quantifying how often a team returns to a previously maintained gaze pattern or *configuration*. These recurrent gaze configurations may signify how team members divide attention and roles, indicative of Individual Visual Attention (IVA), but may also suggest JVA when members share a gaze pattern. Overall, MdRQA captures the consistency in a team's visual attention.

CRQA and MdRQA, are widely utilized in time-series analysis across various fields [2, 34, 69]. In a CPS task, Eloy et al. [17] applied MdRQA on multimodal data including speech, body movements, and skin conductance, and found that higher team-level regularity led to an unpleasant collaboration experience, as it negatively predicted emotional valence. These tools supply a suite of assumption-free nonlinear metrics which we use to analyze the dynamics of visual attention among dyadic and triadic teams.

Recurrence Plots and Metrics. RQA visualizes repeating patterns or regularities in a system using graphical representations known as recurrence plots. A cross-recurrence plot (CRP) visualizes the temporal structures by plotting the recurrence matrix with the gaze trajectories of two individuals on the x and y axes. A point in the matrix is recurrent if the gaze coordinates from the two signals fall within a radius r, measured by some distance metric such as Euclidean distance. The points, forming textures, reveal patterns in the signals' interaction. Recurrent points on the main diagonal represent similar behaviors of the two signals at the same moment - which in our case would mean that the collaborators are looking at the same area simultaneously. As one moves away from the diagonal, the delay between the coupling increases where one individual follows the other at a certain lag. In general, diagonal structures indicate gaze revisitations during the interaction, while horizontal and vertical structures show prolonged viewing of the same area, although it need not have occurred simultaneously.

Recurrence metrics are computed by quantifying a CRP's diagonal and vertical structures, such as recurrence rate, determinism, longest diagonal line, average diagonal line, entropy, laminarity, and trapping-time [63]. Recurrence rate (RR) measures the density of recurrent points, quantifying the similarity of the two gaze signals. Determinism (DET) measures the percentage of recurrent points forming diagonal lines or identical gaze scanpaths. Longest diagonal line (LDL) represents the longest uninterrupted period in sync, and average diagonal line (ADL) the average period in sync. Entropy (ENTR) measures the complexity of coupling, where entropy is low for identical diagonal lengths and high for random diagonal lengths. Laminarity (LAM) measures the percentage of recurrent points forming vertical lines, indicating long durations spent on one target area. Trapping time (TT) is the average duration spent on any specific area, similar to the average vertical line (AVL).

By contrast, MdRQA calculates a symmetric multidimensional recurrence plot (MdRP), combining two or more individuals' gaze trajectories and analyzing their regularity as a single system, resulting in a single time series forming both the x and y axes. Unlike CRP, recurrent points in MdRP denote instances where the system returns to the same gaze *configuration*, regardless of the similarity between each individual's gaze. Beyond recurrence rate, most of the same metrics can be extracted from an MdRP as a CRP. MdRQA therefore uniquely quantifies *team-level* dynamics, while CRQA quantifies *interpersonal* dynamics.

1.1 Related Work

The temporal dimension is especially relevant in complex collaborative environments, where gaze behavior is expected to dynamically adapt to the changing requirements of the task at hand. Accordingly, researchers have leveraged RQA to investigate temporal dynamics in eye gaze during CPS. Devlin et al. [14] found that high-performing dyads in a co-located collaborative task exhibited greater CRQA RR than low-performing dyads, suggesting that sustaining high levels of JVA is beneficial for team performance. In another co-located study, Villamor and Rodrigo [55] used CRQA to analyze gaze data from a pair programming task and found no significant differences among the CRQA metrics for successful and unsuccessful pairs. However, they were able to predict team performance using RR and DET as significant predictors when pretest programming scores were included in the regression model.

Both studies, however, only examined cross-recurrence metrics (RR and DET), hence focusing only on the role of JVA in collaboration and overlooking other important factors such as individual focus, which could imply task-splitting. Additionally, the dyads in both co-located studies worked on separate computers, and in Villamor and Rodrigo [55], verbal communication was discouraged in favor of a chat program. As such, these findings may not be generalizable to naturalistic collaborations without such restrictions.

Villamor et al. [56] later used additional CRQA metrics to identify collaboration patterns in an identical task. Their study demonstrated lower CRQA measures in successful pairs, contradicting previous work [14]. Additionally, they found successful pairs had lower loci and sequence similarity, suggesting a lack of shared visual focus and indicating task-splitting. Limited subject understanding led unsuccessful pairs to fixate on items longer while trying to grasp its meaning, resulting in increased RR and LAM. Additionally, the study's use of chat windows as a communication method (verbal communication was discouraged due to the eye tracker's inability to track gaze when dyads looked away from the screen to communicate) may have affected these results as the windows were in different locations, potentially contributing to lower RR among successful pairs who communicated more. Our work addresses this confound by employing wearable eye trackers.

Gaze-based CRQA and speech data were examined by Schneider et al. [47] during a dyadic co-located collaborative task using wearable eye trackers. They found that JVA positively correlated with interaction quality, but this correlation decreased with increase in participants' expertise. They also found that high coordination (JVA) within a dyad promoted learning gains, but only up to a certain threshold. Our study extends this work by additionally evaluating team-level gaze regularity (measured via MdRQA), which also contributes to a team's likelihood of success.

In a more closely related work, Moulder et al. [36] demonstrated that different RQA-based methods (RQA, CRQA, and MdRQA) on gaze signals show strong predictive capacity for a triadic team's success during a remote CPS task. Vertical ENTR and diagonal ENTR consistently ranked among the top 50% in terms of feature importance for both logistic regression and random forest models, suggesting that increased uncertainty in gaze patterns or exploratory visual behavior, contributed to team success. Vrzakova et al. [59] also examined both CRQA and MdRQA metrics in a block programming task involving remotely-located triads who communicated through a video conferencing platform. They found that gaze-UI regularity predicted a team's negotiation and coordination scores, and gaze-UI coupling predicted the participants' ability to construct a shared knowledge space.

While both studies offer valuable perspectives on the dynamics of *remote* collaboration, they had limitations stemming from the non-naturalistic collaborative environment and data loss from unconstrained head movements when using a screen-based eye tracker. Additionally, their use of categorical RQA with large AOIs may have increased false-positive recurrent points. Our study not only addresses these confounds but also examines collaboration dynamics among dyadic and triadic teams, regardless of task UI changes [59], making it a more generalizable finding in terms of the CPS task and team size. In addition, we examine the relationship between gaze dynamics and subjective traits, adding another dimension to our understanding of CPS.

1.2 Research Questions, Novelty, and Contributions

In this study, we delve into the intricate relationship between a team's visual attention dynamics, their collaborative states, and task performance in order to assess their visual attention-based collaborative strategies. By utilizing metrics derived from CRQA and MdRQA, we analyze gaze coupling and gaze regularity from dyadic and triadic teams engaged in a group programming task. Our research stands out because it examines collaboration dynamics in a naturalistic, co-located environment using wearable eye trackers, enabling participants to engage freely in unconstrained interactions. To our knowledge, this is the first study exploring collaboration dynamics by applying RQA to gaze data in such a context.

We tackle three research questions in this work. First, *how do nonlinear dynamics of visual attention relate to subjective collaborative traits **(RQ1)**?* By examining this connection, our research seeks to illuminate the interplay between objective indicators (characterized by RQA metrics) and subjective facets of productive collaborations, thereby identifying which gaze patterns are positively or negatively associated with them. This understanding will serve as an early indicator of collaboration challenges. Second, *what is the potential of gaze-based RQA metrics in predicting subjective collaborative outcomes? **(RQ2)**?* By identifying these key metrics, we aspire to establish predictive models that inform real-time interventions, ultimately leading to more efficient collaborations. The third and final research question, *in what form do nonlinear patterns of visual attention influence performance in collaborative tasks **(RQ3)**?* Past work has yielded inconsistent findings on the relationship between JVA and collaborative outcomes. By answering this question, we aim to clarify in what aspect does JVA affect collaborative success and whether there are other hidden dynamics at play, thereby providing valuable insights into effective collaboration strategies.

Our contributions are threefold. We advance the understanding of nonlinear gaze dynamics in successful collaborations, identify key gaze-based RQA metrics for predicting subjective collaboration qualities, and provide deeper insights into the dynamic roles that JVA and IVA play at the team-level in promoting successful collaborations.

2 METHOD

2.1 Participants

The study included 44 participants (22 males, 21 females, and 1 non-binary) associated with a university in the United States, including

26 undergraduates and 18 graduate students. Among them, 33 were between 18-24 years old, 9 were between 25-35 years old, and 1 was between 36-46 years old. Participants were compensated $25 per hour for their participation in the study. The IRB-approved study (IRB# 21-0371) lasted a maximum of 2.5 hours, during which participants completed a set of 5 activities in teams of two (n = 13 teams) or three individuals (n = 6 teams). Team size was randomly determined based on session attendance, with groups comprising individuals of mixed genders. Out of 19 teams, 9 teams consisted of unacquainted participants, 7 teams of acquainted participants and 3 triadic teams of 2 acquainted participants.

2.2 Materials

CPS can take many forms, including group programming [62, 68] which is effective in teaching introductory computer science [33, 65]. Block programming tools like MakeCode [11] let users grasp programming concepts while controlling hardware sensors [13, 22, 64]. In our work, MakeCode is used to control a sound sensor which can detect the presence and intensity of noise. The sound sensor is connected to a LED display that connects to the computer, and participants are able to display information from the sound sensor using the LED display. The MakeCode task included two components related to detecting and reporting on sound in the room. The first component was a tutorial activity where participants were taught how to use the MakeCode environment and the corresponding hardware. The second component involved debugging an existing program with the goal of displaying a sad face on the LED display anytime the sound in the room reached a particular threshold. There were two bugs: (a) an unreasonably high threshold value in the block code and (b) an issue with the hardware wiring.

2.3 Experimental Procedure

All participants completed consent forms upon arriving at the space and then went through an eye tracker calibration process with the Tobii Pro Glasses II [53]. Once all participants were calibrated, the eye trackers and additional recording devices (webcam and microphone) started their recordings. Participants completed a brief ice-breaker activity where they shared facts about themselves with the other participants. The team then began a series of 5 co-located group tasks, each followed by a battery of self-report survey measures. The order of the tasks was fixed for all teams, and for the purpose of the present study, we focus on the completion of the MakeCode activity (fourth activity in the series, see Figure 1 (left)). For this task, the researcher introduced the tutorial that the team then completed on how to use the MakeCode interface and hardware. The researcher then presented the team with adjusted hardware (i.e., one of the wires was unplugged out of view of the participants) and the main portion of the task involving debugging the program. Participants were given a maximum of 15 minutes to get the program to work. After 15 minutes, the researcher would stop the participants and prompt them to complete the corresponding self-report surveys.

2.4 Measures

2.4.1 Task Performance Measures. The MakeCode task performance was objectively evaluated based on whether the team resolved the bug in the code (*bug success*) or rewired the hardware to fix the sound sensor issue (*wiring success*). The team received one point for each successful task and an overall score of two if they completed both, i.e., compiled their code so that the hardware depicted a sad face when the volume was above the threshold. If they had no success, they received an overall score of zero.

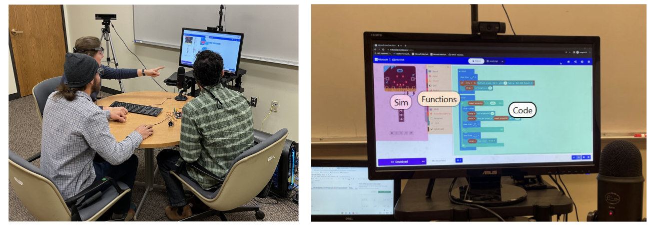

Figure 1: A triad doing the MakeCode task (left), and a snapshot of the MakeCode task used in Assisted Mapping with the AOIs highlighted (right).

2.4.2 Subjective Measures. The subjective assessments included scores on eight self-reported surveys, computed as a team average by taking the mean across every individual for a given team. The surveys include the social cohesion scale [46] to capture perceptions of how well the group is getting along and feelings of belonging; positive interaction [58] to capture perceptions of how well the group was working together towards accomplishing the task; social loafing [31] to capture perceptions of self or others in the group not contributing towards the task; cognitive workload (NASA-TLX [24]) to capture perceptions of the mental demand and success of the task; psychological safety (Team Psychological Safety Scale [16]) for assessing trust in teams; pleasure and arousal (Self-Assessment Manikin [6]) to capture ratings of positive/negative and sleepy/active feelings; and finally, trust [32] to capture perceptions of cognitive and affective trust. These surveys were explicitly selected as they capture the range of social and cognitive processes involved in CPS [52]. Additionally, a programming rating measure was derived from participants' self-reported prior computer programming experience, using a 4-point scale ranging from *"Never computer programmed"* to *"Expert computer programmer"*.

2.4.3 Eye-Tracking Measures. The participants' gaze data are captured from the wearable Tobii Pro Glasses II [53] eye tracker, with a sampling rate of 100Hz. The gaze coordinates are mapped onto a common reference frame to analyze visual attention amongst the team members. We used Tobii Pro Lab's [54] Assisted Mapping feature, which automatically maps gaze data onto still images. We focus our analysis only on the visual attention directed towards the block programming interface, which is the most relevant and informative AOI for our study. Accordingly, all participants' gaze data were mapped onto a snapshot of a computer depicting the MakeCode window and its immediate surroundings (see Figure 1 (right)). Researchers manually corrected incorrect mappings. Furthermore, the mapped gaze coordinates are categorized into one of three primary AOIs: code, functions, and simulation (see Figure 1 (right)). A fourth AOI, labeled "other," is used when the gaze coordinates do not fall within any of the aforementioned three AOIs.

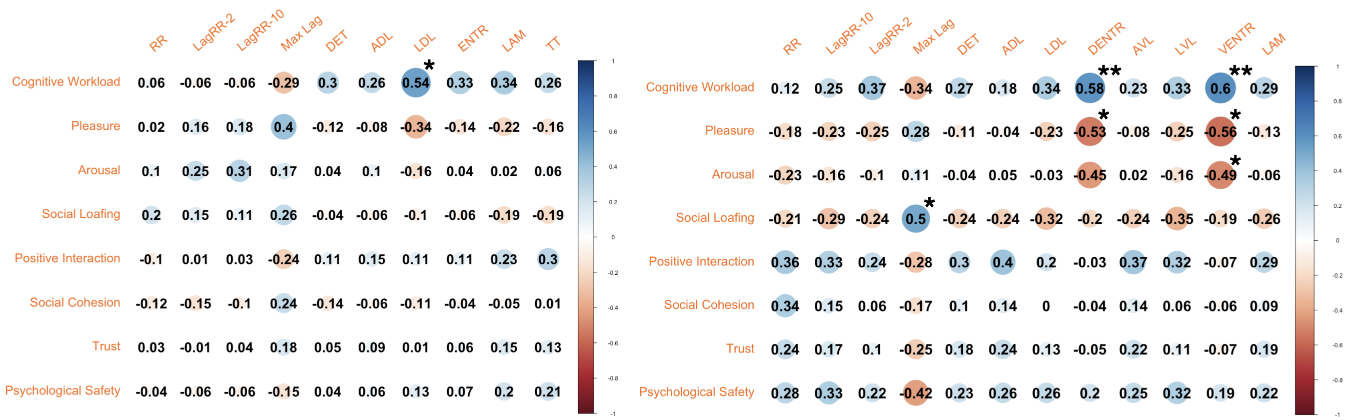

Figure 2: A correlation matrix depicting correlations between the subjective assessments and both CRQA metrics (left) and MdRQA metrics (right); with significant correlations starred (* corresponds to $p \leq 0.05$, and ** corresponds to $p \leq 0.01$).

Data Preprocessing. The raw mapped gaze coordinates were cleaned and downsampled before being subjected to RQA. Due to our desire to maintain ecological validity, head movement was not restricted, resulting in expected data loss. On average, we observed a 29.2% (SD = 10.2) loss of data across teams. Consequently, the timestamps with missing gaze points for one participant (either due to the eye-tracking glasses failing to capture gaze or the participants not viewing the MakeCode window) were eliminated from the time series of all other participants in the team. This was done to ensure consistency in the time series while computing RQA, and therefore we have included the length of the time series as a covariate in our regression models. The cleaned gaze data was then downsampled from 100Hz to 10Hz to reduce noise in the raw signal and make it computationally inexpensive for RQA. We used raw gaze data rather than fixations to reduce data loss, following [59].

RQA Metrics and Parameters. In addition to the RQA metrics explained in Section 1, we also compute metrics such as Maximal Lag (Max Lag) and Lag Recurrence Rates (LagRR). Max Lag is the index of a diagonal (other than the main diagonal), with the highest density of recurrent points, whereas LagRR is a measure of the density of recurrent points within a specified lag window from the main diagonal. We compute LagRR for a 2-second (LagRR-2) and a 10-second window (LagRR-10) while operationalizing CRQA LagRR for the 2-second window to denote JVA, i.e., the gaze-coupling between two collaborators within a 2-second delay, based on prior work [47] and since past research has noted that 2 seconds is the delay between a listener following a speaker's gaze [44]. We chose the 10-second window based on the eye tracker sampling rate, per recommendations by previous work [12].

We computed continuous input-based CRQA and MdRQA metrics in R using the *crqa* package [10] and the *mdrqa* package [60], respectively. We note that the *mdrqa* package additionally outputs a longest vertical line (LVL) metric and distinguishes ENTR as two individual measures: one for the diagonal (DENTR) and one for vertical lines (VENTR). We set the threshold radius for CRQA to be 35 pixels on the snapshot image, based on our expert judgment to balance between accounting for eye tracker measurement error while not categorizing false positive recurrent points. The threshold radius for MdRQA was fixed at 0.33 so that the average RR for all teams was ~ 5%, as is the recommended practice for comparative recurrence analyses [60, 63]. Both CRQA and MdRQA employed Euclidean distance as the distance metric, without any additional data embedding, following [25]. Lastly, we compute the CRQA metrics for triads as the pairwise metrics between all three team members and then average them for each team.

3 RESULTS

We delve into three research questions pertaining to the dynamics of visual attention in CPS, and present our findings accordingly.

RQ1: How do nonlinear dynamics of visual attention relate to subjective collaborative traits? To begin exploring the links between the nonlinear gaze dynamics and the outcome measures, we computed pairwise Pearson correlations between the subjective assessments listed in Section 2.4.2 and the CRQA/MdRQA metrics. These correlations are visualized as a correlation matrix in Figure 2, with CRQA correlations on the left and MdRQA correlations on the right. A significant positive correlation was observed between the cognitive workload measure and CRQA LDL (r = .54, $p < 0.05$). No other significant correlations were observed between the CRQA metrics and the subjective traits. With respect to the team-level MdRQA metrics, reported cognitive workload showed significant positive correlations with DENTR (r = .58, $p < 0.01$) and VENTR (r = .6, $p < 0.01$). Additionally, the pleasure metric negatively correlated with DENTR (r = -.53, $p < 0.05$) and VENTR (r = -.56, $p < 0.05$), while the arousal metric negatively correlated with VENTR (r = -.49, $p < 0.05$). Notably, social loafing was positively correlated with Max Lag (r = .5, $p < 0.05$).

RQ2: What is the potential of gaze-based RQA metrics in predicting subjective collaborative outcomes? To comprehensively test the effects of various RQA metrics, we performed a model selection procedure to identify the best fit linear regression model for each subjective outcome based on the Bayesian Information Criteria (BIC). BIC rates the estimated models based on the amount of explained variance in the dependent variable while penalizing complexity (number of additional parameters) in a model [48]. For

Table 1: Regression Models Predicting Subjective Assessments of Collaborative Traits

Subjective Trait	Metric 1	Metric 2	Metric 3	Covariates	
Cognitive Workload Effect Size (*p*)	CRQA LagRR-10 -0.97 (***0.007***)	CRQA LDL 1.70 (***0.004***)	MdRQA LVL -0.94 (***0.013***)	Team Size 0.19 (*0.554*)	Length 0.31 (*0.126*)
Pleasure Effect Size (*p*)	MdRQA DET 2.77 (***0.003***)	CRQA LAM -1.70 (***0.005***)	MdRQA DENTR -2.67 (***0.006***)	Team Size -0.89 (*0.063*)	Length 0.88 (***0.043***)
Arousal Effect Size (*p*)	CRQA LagRR-10 0.31 (*0.238*)			Team Size 0.52 (*0.286*)	Length -0.40 (*0.089*)
Social Loafing Effect Size (*p*)	MdRQA Max Lag 0.56 (***0.045***)			Team Size 0.44 (*0.408*)	Length 0.07 (*0.785*)
Positive Interaction Effect Size (*p*)	MdRQA LagRR-10 0.7 (***0.037***)			Team Size 0.32 (*0.563*)	Length -0.56 (***0.050***)
Social Cohesion Effect Size (*p*)	MdRQA DET 1.20 (***0.019***)	CRQA ADL 4.33 (***0.030***)	CRQA DET -5.68 (***0.017***)	Team Size 0.23 (*0.646*)	Length -0.22 (*0.379*)
Trust Effect Size (*p*)	MdRQA VENTR 1.1 (*0.114*)			Team Size 0.71 (*0.329*)	Length -1.14 (*0.057*)
Psychological Safety Effect Size (*p*)	MdRQA Max Lag -0.44 (*0.117*)			Team Size -0.19 (*0.721*)	Length -0.03 (*0.899*)

each outcome variable, we fitted models of varying complexity including each combination of one, two, or three recurrence metrics (CRQA or MdRQA) while controlling for the effects of team size and time series length as common covariates across all models. To identify the gaze dynamics that collectively contribute to collaborative outcomes, for each dependent measure we selected the model with the lowest BIC and a BIC difference of no less than 3 from all other models [29]. We performed the Benjamini-Hochberg false discovery rate procedure to correct for multiple comparisons made during model selection [5]. Final models, presented with adjusted p-values in Table 1, had all variables individually z-scored such that the reported effect sizes are in terms of standard deviations.

Reported cognitive workload was best explained by a 3-metric model with CRQA LagRR-10 ($B = -0.97, p = 0.007$), MdRQA LVL ($B = -0.94, p = 0.013$) in the negative direction and CRQA LDL ($B = 1.70, p = 0.004$) in the positive direction ($R^2 = 0.789; RMSE = 0.460$). Perceived pleasure was best fit by a combination of MdRQA DET ($B = 2.77, p < 0.003$), MdRQA DENTR ($B = -2.67, p = 0.006$), and CRQA LAM ($B = -1.70, p = 0.005$) ($R^2 = 0.786; RMSE = 0.477$). Similarly, social cohesion was best fit to MdRQA DET ($B = 1.20, p < 0.019$), CRQA ADL ($B = 4.33, p < 0.030$), and CRQA DET ($B = -5.68, p < 0.017$), all showing significant effects ($R^2 = 0.581; RMSE = 0.676$). Single-metric models were chosen for arousal (CRQA LagRR-10: $B = 0.31, p < 0.238$), trust (MdRQA VENTR: $B = 1.10, p < 0.114$), and psychological safety (MdRQA Max Lag: $B = -0.44, p < 0.117$), but none displayed significant effects. For social loafing, MdRQA Max Lag displayed a significant positive effect ($B = 0.56, p < 0.045$) ($R^2 = 0.285; RMSE = 0.874$). Lastly, positive interaction was best modeled by MdRQA LagRR-10 ($B = 0.70, p < 0.037$) ($R^2 = 0.340; RMSE = 0.860$).

RQ3: In what form do nonlinear patterns of visual attention influence performance in collaborative tasks? To identify the type of gaze dynamic with the most impact on collaborative success, we looked at the potential of all CRQA and MdRQA metrics in predicting the objective task performance, which was coded as 0, 1, or 2 based on the number of solved bugs. To appropriately model this 3-level outcome, ordinal logistic regression was performed to assess which recurrence measures best predict the odds of team success. The same model selection procedure as above was performed, allowing for 1-3 RQA metrics to be included with team size, time series length, and average programming rating measure as covariates. Odds of achieving partial or complete success versus no success in the task was best modeled by a single metric, MdRQA LVL: odds ratio = 173.37 ($p < 0.001$), meaning that a team is 173 times more likely to obtain partial or complete success as MdRQA LVL increases by one standard deviation. Notably, a 0.1-second increase in the LVL measure was associated with a 26% higher likelihood of success, highlighting a major potential relationship with team performance.

4 DISCUSSION

The present study examined the relationship between nonlinear gaze dynamics and collaborative traits, evaluated the predictive capabilities of nonlinear gaze metrics, and delved into the specific ways in which JVA and IVA contribute to collaborative success.

A significant positive correlation between cognitive workload and CRQA LDL suggests that increased workload is associated with sustained gaze coupling periods or JVA, highlighting the need for shared understanding during demanding tasks [42]. The positive correlations with MdRQA VENTR and DENTR imply that team members exhibit more exploratory visual behavior during challenging tasks, possibly to gather more information and generate new ideas. Conversely, the negative association of CRQA LagRR-10 and MdRQA LVL in predicting workload indicates that increased gaze coordination and sustained team-level regularity suggest better cognitive resource management among team members, ultimately reducing overall workload. This multifaceted relationship provides valuable insights for optimizing team performance, emphasizing the importance of striking a balance between JVA, which fosters shared understanding, and IVA for effective task allocation and workload distribution.

Moreover, we discovered a positive correlation between social loafing and MdRQA Max Lag, which also serves as the best predictor for this trait. MdRQA Max Lag represents the time delay between when the collaborators' gaze patterns exhibit maximum regularity or return to a previously established gaze configuration. This configuration could signify engagement in an individual task or a joint coordinated effort with team members. The finding suggests a possible association that an increased delay corresponds to a higher likelihood of loafing behavior. Therefore, teams that can swiftly adapt and re-establish a shared or individual focus could potentially minimize loafing behavior, leading to improved collaboration outcomes. This novel insight can be harnessed to develop monitoring strategies and real-time interventions for mitigating social loafing, ultimately enhancing team performance and productivity.

Arousal exhibited a negative correlation with MdRQA DENTR, while pleasure showed negative correlations with both MdRQA DENTR and VENTR, indicating that as teams display more exploratory visual behavior (which, as mentioned earlier, signifies high workload), they experience reduced positive emotions and arousal. This observation could be attributed to the possibility that positive emotions and arousal primarily stem from more coordinated efforts. Furthermore, along with MdRQA DENTR and CRQA LAM negatively predicting pleasure, MdRQA DET displayed a positive association, suggesting that teams with greater coordinated and predictable visual behavior will likely experience heightened

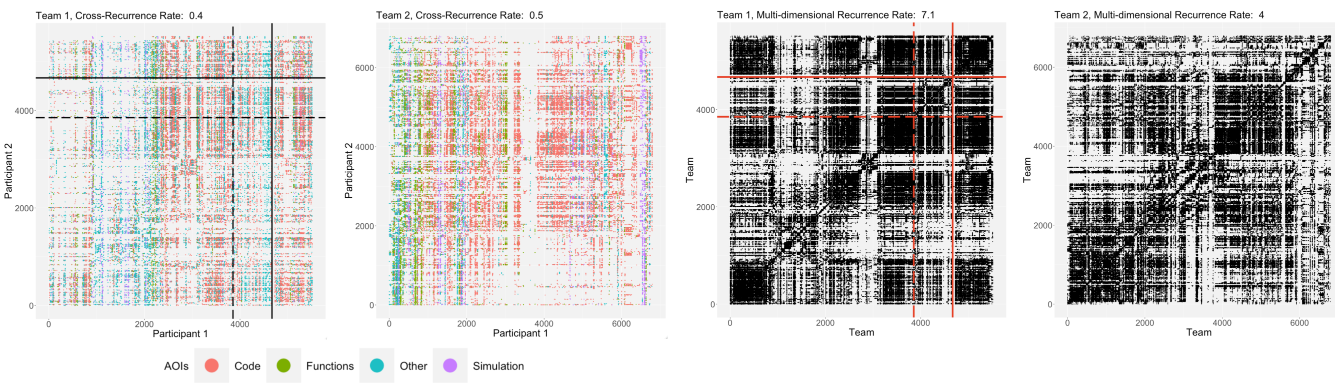

Figure 3: AOI-based color-coded cross-recurrence plots depicting gaze coupling (left) and Multi-dimensional recurrence plots depicting team-level gaze regularity (right) of a high-performing (Team 1) and a low-performing team (Team 2). The times when team 1 achieved wiring and bug success are denoted by the dashed and solid lines, respectively.

pleasure. Conversely, as CRQA LAM reflects collaborators remaining focused (or stuck [56]) on a specific region, an increase in this metric was associated with diminished pleasure. These findings suggest a link between coordinated attention allocation and positive emotions during collaborative tasks.

A single metric, MdRQA LagRR-10, best predicted positive interaction among collaborators, implying that teams maintaining high gaze regularity or consistent individual or joint gaze behavior within a 10-second lag may have better quality interactions. Social cohesion, on the other hand, was best predicted by a combination of MdRQA DET, CRQA ADL, and CRQA DET metrics. MdRQA DET represents the presence of predictable team regularity, while CRQA ADL signifies the stability of gaze coupling among team members. Their positive associations suggest that teams with more stable gaze coupling and predictable gaze regularity may likely experience better cohesion. The seemingly counter-intuitive combination of positive MdRQA DET and negative CRQA DET reveals a unique dynamic: predictability of team-level configurations, rather than JVA, underlies social cohesion. As MdRQA accounts for both JVA and IVA, this suggests that team-level dynamics with IVA play a more significant role in social cohesion, than JVA.

Stability in CPS. Lastly, ordinal logistic regression supported the predictive power of team-level gaze dynamics, specifically MdRQA LVL, on task performance. The measure of LVL in an MdRP captures the longest span in which a team remained in the same *configuration*, essentially reflecting the team's stability. To form these vertical structures, team members can have close or distant gaze points, provided they collectively show minimal movement from their positions. In the context of collaborative processes, this finding is likely indicative of establishing roles or task-splitting among team members, showcasing team-level coordination and IVA. It can also represent sustained periods of joint attention, where team members align their attention and address complex aspects of the task. Fitted odds ratios reveal that for every 0.1 seconds a team remains in their longest configuration, they are 26% more likely to obtain partial or complete success. While our results reaffirm the role of JVA in collaboration, they also emphasize the significance of maintaining a stable gaze configuration, accommodating either IVA or JVA. Particularly when compared to brief or fragmented manifestations of solely JVA reported in prior work [47, 55]. Beyond gaze, verbal patterns may provide additional cues for role assignment, task clarification, and coordination. Teams with lower entropy (reflecting stability) in their communication sequences during CPS phase transitions have demonstrated better task performance [66].

This stability quantifies the degree of sustained investment by all team members and depicts a well-functioning team with a clear understanding of their objectives and roles. Team members may have developed trust and communication patterns that enable them to effectively allocate their attention and resources to maximize their collaborative success [9]. A team with high JVA collaboratively addressing a task may show sustained MdRQA LVL, reflecting their cooperation. Conversely, another team focusing individually before regrouping might exhibit high LVL due to prolonged IVA, indicating their distinct approach.

The team-level nature of MdRQA means that a team's specific configuration doesn't matter, only the time spent in it. These results can be interpreted in light of multiple findings in the team cognition and CPS literature that study teams as dynamical systems. Research has revealed that teams' ability to adaptively and efficiently transition through stable configurations while minimizing transition time is associated with increased team effectiveness [19, 20]. Moreover, studies have characterized transitions between stable configurations as times of team uncertainty, with effective teams minimizing time spent in transitional, or unstable configurations [50]. Within CPS, stable configurations are considered to represent distinct phases in the CPS process that contribute to team success [66]. Given these findings, metrics like MdRQA LVL are highly adaptable to each team's unique strategy and composition, imposing no assumptions or constraints on the form of collaboration while capturing its configural stability. This maintains generalizability across tasks and individuals, supporting MdRQA's viability for monitoring successful co-located collaborations.

Qualitative Anecdotes to Visualize Findings. We present anecdotal evidence comparing a successful dyad (team 1 = *wiring and bug success*) and an unsuccessful dyad (team 2 = *no success*) by visualizing their gaze behavior via CRPs, MdRPs, and gaze-based heat maps. Figure 3 shows a CRP's recurrent points color-coded by the jointly focused AOI on the left and MdRPs on the right, with dashed and solid lines marking *wiring and bug success*, respectively. Though team 2 displayed more instances of gaze-coupling and JVA (RR = 0.5, LagRR-2 = 0.62) than team 1 (RR = 0.4, LagRR-2 = 0.47), they failed the task. Despite higher JVA, team 2's lack of success may stem from their high CRQA LAM (69.4%) and TT (3.3), suggesting prolonged single-region focus due to limited subject understanding [26]. In contrast, team 1 had lower LAM (44.2%) and TT (2.6). This behavior, called 'staring but not seeing,' is also noted in Villamor et al. [56]. Remarkably, both teams had almost similar programming ratings (team 1 = 1; team 2 = 0.5).

Interestingly, an empty gap in team 1's CRP below the intersection of the solid lines reveals no coupling right before solving the bug, suggesting a greater emphasis on IVA during that critical moment. Moreover, team 1 exhibited exploratory visual attention, including non-AOI regions (blue textures), before achieving bug success. Figure 4 also confirms team 1's more varied gaze density distribution, reflecting an exploratory approach, unlike team 2. This finding underscores the need for a balanced and flexible approach, incorporating both gaze coupling and individual exploration in visual attention during collaboration.

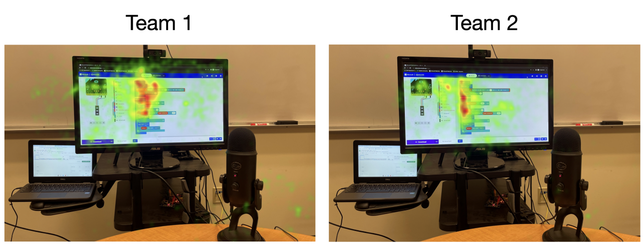

Figure 4: Heat-maps of a high-performing (Team 1; bug and wiring success) and a low-performing team (Team 2; no success) depicting their visual attention throughout the task.

Furthermore, team 1 displayed higher overall gaze regularity (RR = 7.1) than team 2 (RR = 4). This is evident in the MdRPs, where team 1 displays denser textures along the diagonal and towards the task's end, as well as prior to resolving the hardware wiring issue. Team 2 showed regularity around the 4000^{th} mark, but recurrent point density diminished towards the task's end. These texture characteristics suggest that maintaining gaze configuration is crucial for task success, further corroborating our findings.

Team 1, who knew each other beforehand, exhibited higher trust and social cohesion (4.8 and 5, respectively) than team 2 (4.3 and 3.6, respectively), who met during the study, indicating more familiarity and bonding in team 1. They also had a lower workload rating (7.5) compared to team 2 (14.1). This interpersonal trust and rapport may have enhanced coordination and understanding of the task objectives for team 1, leading to more stable gaze configurations and workload distribution. These findings additionally underscore the vital role of team-level regularity and confirm MdRQA and its metrics as valuable tools for evaluating collaborative dynamics.

4.1 Applications

This work's potential applications can impact various fields where effective collaboration is vital. In education, educators can use these findings to create targeted training programs that teach students effective coordination to either jointly solve tasks or focus on individual responsibilities, while stressing the importance of stability and flexibility in their approach. In professional settings, managers can leverage these insights, such as optimized workload distribution, to enhance team performance. This can be achieved through team-building exercises that focus on establishing trust and social cohesion. Additionally, this research could lead to real-time monitoring systems that detect potential issues like excessive team workloads and social loafing, allowing for timely interventions to maintain high-quality collaborative experiences. It can also guide when to avoid interventions, such as during sustained periods of JVA/IVA in teams. Exploring multimodal communication channels (e.g., speech and gestures) alongside gaze can offer a deeper understanding of collaboration dynamics and inform more effective intervention designs. Beyond human collaboration, these findings could also inform the development of intelligent collaborative technologies that better understand and adapt to human behavior.

4.2 Limitations and Future Work

This study had certain limitations that future research can address. First, our sample size was moderate, with only 19 teams analyzed, which may constrain the robustness of our statistical analysis. Nonetheless, our qualitative observations further supported our findings. Second, our dataset includes both dyadic and triadic teams, but we addressed this limitation by adding team size as a covariate in our analysis. Though team size didn't significantly affect results, suggesting our approach might be generalizable across team sizes, further research is required to confirm this. Moreover, the exploratory nature of our study, examining the interactions between various RQA metrics and multiple subjective and objective outcomes, may constrain the generalizability of our findings. Though it is worth noting that this is the first work to explore collaboration dynamics through eye gaze in a naturalistic context. Further research can benefit from a more targeted approach. Finally, another limitation is that we focused on a single task context, block programming in MakeCode. In future work, we'll study gaze behavior in various CPS tasks alongside data streams like speech to see if MakeCode task patterns apply to other CPS contexts.

5 CONCLUSION

In conclusion, this study has offered valuable insights into the complex function of nonlinear visual attention dynamics, and the importance of team-level regularity, specifically the regularity of JVA and IVA, in successful collaborations. By understanding these intricate visual attention strategies, we can better optimize teamwork across various domains and facilitate more effective collaboration.

ACKNOWLEDGMENTS

This research was supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019805. The opinions expressed are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

[1] Angelina Abitino, Samuel L Pugh, Candace E Peacock, and Sidney K D'Mello. 2022. Eye to Eye: Gaze Patterns Predict Remote Collaborative Problem Solving Behaviors in Triads. *Artificial Intelligence in Education: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part I*, 378–389. https://doi.org/10.1007/978-3-031-11644-5_31

[2] U Rajendra Acharya, S Vinitha Sree, Subhagata Chattopadhyay, Wenwei Yu, and Peng Chuan Alvin Ang. 2011. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. *International journal of neural systems* 21, 03 (2011), 199–211.

[3] Oya Aran and Daniel Gatica-Perez. 2013. One of a Kind: Inferring Personality Impressions in Meetings. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, Australia) *(ICMI '13)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/2522848.2522859

[4] Daniel Belenky, Michael Ringenberg, Jennifer Olsen, Vincent Aleven, and Nikol Rummel. 2014. Using Dual Eye-Tracking to Evaluate Students' Collaboration with an Intelligent Tutoring System for Elementary-Level Fractions. *Grantee Submission* (2014).

[5] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

[6] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.

[7] Susan E Brennan, Xin Chen, Christopher A Dickinson, Mark B Neider, and Gregory J Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106, 3 (2008), 1465–1477.

[8] Mehmet Celepkolu and Kristy Elizabeth Boyer. 2018. Predicting Student Performance Based on Eye Gaze During Collaborative Problem Solving. *Proceedings of the Group Interaction Frontiers in Technology, GIFT 2018.* https://doi.org/10.1145/3279981.3279991

[9] Xusen Cheng, Shixuan Fu, Jianshan Sun, Yajing Han, Jia Shen, and Alex Zarifis. 2016. Investigating individual trust in semi-virtual collaboration of multicultural and unicultural teams. *Computers in Human Behavior* 62 (2016), 267–276.

[10] Moreno I Coco and Rick Dale. 2014. Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in psychology* 5 (2014), 510.

[11] Microsoft Corporation. 2023. *Microsoft MakeCode.* Retrieved May 2, 2023 from https://www.microsoft.com/en-us/makecode

[12] Rick Dale, Anne S Warlaumont, and Daniel C Richardson. 2011. Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos* 21, 04 (2011), 1153–1161.

[13] James Devine, Joe Finney, Peli de Halleux, Michał Moskal, Thomas Ball, and Steve Hodges. 2018. MakeCode and CODAL: intuitive and efficient embedded systems programming for education. *ACM SIGPLAN Notices* 53, 6 (2018), 19–30.

[14] Shannon Patricia Devlin, Jake Ryan Flynn, and Sara Lu Riggs. 2020. How Shared Visual Attention Patterns of Pairs Unfold Over Time when Workload Changes. In *ACM Symposium on Eye Tracking Research and Applications*. 1–5.

[15] Vanessa Urch Druskat and D Christopher Kayes. 2000. Learning versus performance in short-term project teams. *Small group research* 31, 3 (2000), 328–353.

[16] Amy Edmondson. 1999. Psychological safety and learning behavior in work teams. *Administrative science quarterly* 44, 2 (1999), 350–383.

[17] Lucca Eloy, Angela E.B. Stewart, Mary Jean Amon, Caroline Reinhardt, Amanda Michaels, Chen Sun, Valerie Shute, Nicholas D. Duran, and Sidney D'Mello. 2019. Modeling Team-Level Multimodal Dynamics during Multiparty Collaboration. In *2019 International Conference on Multimodal Interaction* (Suzhou, China) *(ICMI '19)*. Association for Computing Machinery, New York, NY, USA, 244–258. https://doi.org/10.1145/3340555.3353748

[18] Chris D Frith and Uta Frith. 2007. Social cognition in humans. *Current biology* 17, 16 (2007), R724–R732.

[19] Jamie C Gorman, Mustafa Demir, Nancy J Cooke, and David A Grimm. 2019. Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: A layered dynamics approach. *Ergonomics* 62, 5 (2019), 629–643.

[20] Jamie C Gorman, David A Grimm, Ronald H Stevens, Trysha Galloway, Ann M Willemsen-Dunlap, and Donald J Halpin. 2020. Measuring real-time team cognition during team training. *Human factors* 62, 5 (2020), 825–860.

[21] Arthur C Graesser, Stephen M Fiore, Samuel Greiff, Jessica Andrews-Todd, Peter W Foltz, and Friedrich W Hesse. 2018. Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest* 19, 2 (2018), 59–92.

[22] Shuchi Grover and Satabdi Basu. 2017. Measuring student learning in introductory block-based programming: Examining misconceptions of loops, variables, and boolean logic. In *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education*. 267–272.

[23] Julie L Harrison, Jamie C Gorman, and Michael J Crites. 2022. Visuomotor Lag and the Intermanual Speed Advantage. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 66. SAGE Publications Sage CA: Los Angeles, CA, 893–897.

[24] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[25] Joseph S Iwanski and Elizabeth Bradley. 1998. Recurrence plots of experimental data: To embed or not to embed? *Chaos: An Interdisciplinary Journal of Nonlinear Science* 8, 4 (1998), 861–871.

[26] Robert JK Jacob and Keith S Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*. Elsevier, 573–605.

[27] Patrick Jermann, Dejana Mullins, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2011. Collaborative gaze footprints: Correlates of interaction quality. (2011).

[28] Patrick Jermann and Marc-Antoine Nüssli. 2012. Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 1125–1134.

[29] Robert E Kass and Adrian E Raftery. 1995. Bayes factors. *Journal of the american statistical association* 90, 430 (1995), 773–795.

[30] John M Levine and Richard L Moreland. 2008. *Small groups: key readings*. Psychology Press.

[31] Lisa Linnenbrink-Garcia, Toni Kempler Rogat, and Kristin LK Koskey. 2011. Affect and engagement during small group instruction. *Contemporary Educational Psychology* 36, 1 (2011), 13–24.

[32] Daniel J McAllister. 1995. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal* 38, 1 (1995), 24–59.

[33] Charlie McDowell, Linda Werner, Heather E Bullock, and Julian Fernald. 2006. Pair programming improves student retention, confidence, and program quality. *Commun. ACM* 49, 8 (2006), 90–95.

[34] Claire Morana, Sofiane Ramdani, Stéphane Perrey, and Alain Varray. 2009. Recurrence quantification analysis of surface electromyographic signal: sensitivity to potentiation and neuromuscular fatigue. *Journal of Neuroscience Methods* 177, 1 (2009), 73–79.

[35] Frederick P Morgeson, Matthew H Reider, and Michael A Campion. 2005. Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel psychology* 58, 3 (2005), 583–611.

[36] Robert Moulder, Brandon Booth, Angelina Abitino, and Sidney D'Mello. 2023. Recurrence Quantification Analysis of Eye Gaze Dynamics During Team Collaboration. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (Arlington, TX, USA) *(LAK2023)*. Association for Computing Machinery, New York, NY, USA, 430–440. https://doi.org/10.1145/3576050.3576113

[37] Robert G. Moulder, Nicholas D. Duran, and Sidney K. D'Mello. 2022. Assessing Multimodal Dynamics in Multi-Party Collaborative Interactions with Multi-Level Vector Autoregression. In *Proceedings of the 2022 International Conference on Multimodal Interaction* (Bengaluru, India) *(ICMI '22)*. Association for Computing Machinery, New York, NY, USA, 615–625. https://doi.org/10.1145/3536221.3556595

[38] Marc-Antoine Nüssli. 2011. *Dual eye-tracking methods for the study of remote collaborative problem solving*. Technical Report. EPFL.

[39] Jennifer K. Olsen, Kshitij Sharma, Nikol Rummel, and Vincent Aleven. 2020. Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology* 51 (2020). Issue 5. https://doi.org/10.1111/bjet.12982

[40] Sami Pietinen, Roman Bednarik, Tatiana Glotova, Vesa Tenhunen, and Markku Tukiainen. 2008. A method to study visual attention aspects of collaboration: eye-tracking pair programmers simultaneously. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. 39–42.

[41] Sami Pietinen, Roman Bednarik, and Markku Tukiainen. 2010. Shared visual attention in collaborative programming: a descriptive analysis. In *proceedings of the 2010 ICSE workshop on cooperative and human aspects of software engineering*. 21–24.

[42] Johanna Pöysä-Tarhonen, Nafisa Awwal, Päivi Häkkinen, and Suzanne Otieno. 2021. Joint attention behaviour in remote collaborative problem solving: exploring different attentional levels in dyadic interaction. *Research and Practice in Technology Enhanced Learning* 16, 1 (2021), 11.

[43] Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. 2021. Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning Analytics - Can We Go the Whole Nine Yards? *IEEE Transactions on Learning Technologies* 14 (2021). Issue 3. https://doi.org/10.1109/TLT.2021.3097766

[44] Daniel C Richardson and Rick Dale. 2005. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science* 29, 6 (2005), 1045–1060.

[45] Jeremy Roschelle and Stephanie D. Teasley. 1995. *The Construction of Shared Knowledge in Collaborative Problem Solving*. https://doi.org/10.1007/978-3-642-85098-1_5

[46] Leisa D Sargent and Christina Sue-Chan. 2001. Does diversity affect group efficacy? The intervening role of cohesion and task interdependence. *Small group research* 32, 4 (2001), 426–450.

[47] Bertrand Schneider, Kshitij Sharma, Sebastien Cuendet, Guillaume Zufferey, Pierre Dillenbourg, and Roy Pea. 2018. Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning* 13, 3 (2018), 241–261.

[48] Gideon Schwarz. 1978. Estimating the dimension of a model. *The annals of statistics* (1978), 461–464.

[49] Lisa J Stephenson, S Gareth Edwards, and Andrew P Bayliss. 2021. From gaze perception to social cognition: The shared-attention system. *Perspectives on Psychological Science* 16, 3 (2021), 553–576.

[50] Ronald Stevens, Trysha Galloway, Donald Halpin, and Ann Willemsen-Dunlap. 2016. Healthcare teams neurodynamically reorganize when resolving uncertainty. *Entropy* 18, 12 (2016), 427.

[51] Shree Krishna Subburaj, Angela E.B. Stewart, Arjun Ramesh Rao, and Sidney K. D'Mello. 2020. Multimodal, Multiparty Modeling of Collaborative Problem Solving Performance. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (Virtual Event, Netherlands) *(ICMI '20)*. Association for Computing Machinery, New York, NY, USA, 423–432. https://doi.org/10.1145/3382507.3418877

[52] Chen Sun, Valerie J Shute, Angela Stewart, Jade Yonehiro, Nicholas Duran, and Sidney D'Mello. 2020. Towards a generalized competency model of collaborative problem solving. *Computers & Education* 143 (2020), 103672.

[53] Tobii Pro AB. 2014. Tobii Pro Glasses 2 eye tracker. Apparatus. Retrieved May 2, 2023 from http://www.tobiipro.com/

[54] Tobii Pro AB. 2022. Tobii Pro Lab. Computer software. Retrieved May 2, 2023 from http://www.tobiipro.com/

[55] Maureen Villamor and Ma Mercedes Rodrigo. 2018. Predicting successful collaboration in a pair programming eye tracking experiment. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. 263–268.

[56] Maureen M Villamor, Ma Rodrigo, and T Mercedes. 2019. Gaze collaboration patterns of successful and unsuccessful programming pairs using cross-recurrence quantification analysis. *Research and Practice in Technology Enhanced Learning* 14, 1 (2019), 1–22.

[57] Maureen M Villamor and Mercedes T Rodrigo. 2017. Characterizing Collaboration in the Pair Program Tracing and Debugging Eye-Tracking Experiment: A Preliminary Analysis. *International Educational Data Mining Society* (2017).

[58] Astrid JSF Visschers-Pleijers, Diana HJM Dolmans, Ineke HAP Wolfhagen, and Cees PM van der Vleuten. 2005. Development and validation of a questionnaire to identify learning-oriented group interactions in PBL. *Medical teacher* 27, 4 (2005), 375–381.

[59] Hana Vrzakova, Mary Jean Amon, Angela EB Stewart, and Sidney K D'Mello. 2019. Dynamics of visual attention in multiparty collaborative problem solving using multidimensional recurrence quantification analysis. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[60] Sebastian Wallot and Giuseppe Leonardi. 2018. Analyzing multivariate dynamics using cross-recurrence quantification analysis (crqa), diagonal-cross-recurrence profiles (dcrp), and multidimensional recurrence quantification analysis (mdrqa)–a tutorial in r. *Frontiers in psychology* 9 (2018), 2232.

[61] Sebastian Wallot, Andreas Roepstorff, and Dan Mønster. 2016. Multidimensional Recurrence Quantification Analysis (MdRQA) for the analysis of multidimensional time-series: A software implementation in MATLAB and its application to group-level data in joint action. *Frontiers in psychology* (2016), 1835.

[62] Noreen M Webb, Philip Ender, and Scott Lewis. 1986. Problem-solving strategies and group processes in small groups learning computer programming. *American Educational Research Journal* 23, 2 (1986), 243–261.

[63] Charles L Webber Jr and Joseph P Zbilut. 2005. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences* 94, 2005 (2005), 26–94.

[64] David Weintrop. 2019. Block-based programming in computer science education. *Commun. ACM* 62, 8 (2019), 22–25.

[65] Laurie Williams, Eric Wiebe, Kai Yang, Miriam Ferzli, and Carol Miller. 2002. In support of pair programming in the introductory computer science course. *Computer Science Education* 12, 3 (2002), 197–212.

[66] Travis J Wiltshire, Jonathan E Butner, and Stephen M Fiore. 2018. Problem-solving phase transitions during team collaboration. *Cognitive science* 42, 1 (2018), 129–167.

[67] Xiangyu Ying, Huanhuan Li, Shan Jiang, Fei Peng, and Zhongxin Lin. 2014. Group laziness: The effect of social loafing on group performance. *Social Behavior and Personality: an international journal* 42, 3 (2014), 465–471.

[68] Z Zacharia, Danielle Boulden, Jessica Vandenberg, Jennifer Tsan, Collin Lynch, Eric Wiebe, and K Boyer. 2019. Collaborative talk across two pair-programming configurations. In *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019*, Vol. 1.

[69] Joseph P Zbilut, Nitza Thomasson, and Charles L Webber. 2002. Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. *Medical engineering & physics* 24, 1 (2002), 53–60.