Dynamics of Visual Attention in Multiparty Collaborative Problem Solving using Multidimensional Recurrence Quantification Analysis

Hana Vrzakova
Institute of Cognitive Science
University of Colorado Boulder
Boulder, Colorado
hana.vrzakova@colorado.edu

Mary Jean Amon
Institute of Cognitive Science
University of Colorado Boulder
Boulder, Colorado
mary.amon@colorado.edu

Angela E. B. Stewart
Institute of Cognitive Science
University of Colorado Boulder
Boulder, Colorado
angela.stewart@colorado.edu

Sidney K. D'Mello
Institute of Cognitive Science
University of Colorado Boulder
Boulder, Colorado
sidney.dmello@colorado.edu

ABSTRACT

Multiparty collaborative problem solving–an increasingly important context in the 21st century workforce–suffers from a degradation of social and behavioral signals when attempted remotely, resulting in suboptimal outcomes. We investigate teams' multidimensional patterns of visual attention during a collaborative problem-solving task with an eye for leveraging insights to improve collaborative interfaces. Fifty-seven novices (forming 19 triads) engaged in a challenging programming task (Minecraft Hour of Code) using videoconferencing software with screen sharing. To discover patterns of individual-level *gaze-UI coupling* (coordination of a teammate's attention with respect to changes in the user interface) and team-level *gaze-UI regularity* (dynamics of teams' collective attention in context with changes in the user interface), we applied cross- and multidimensional recurrence quantification analyses, respectively. Individuals' eye gaze was significantly coupled with the ongoing screen activity whereas teams displayed significant patterns of gaze regularity, suggesting repetitive patterns in teams' attention. These measures predicted expert-coded collaborative processes of constructing shared knowledge and negotiation and coordination (but not maintaining team function) and correlated with task score ($r = .425$). They also predicted individually assessed subjective perceptions of team performance and the collaboration process, but not individual's learning or team's task scores. We discuss implications of our findings for the design of intelligent collaborative interfaces.

CCS CONCEPTS

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

KEYWORDS

CSCL, CSCW, Eye-Tracking, MdRQA, Regularity, Synchrony

1 INTRODUCTION

"Look here!"—"Where? In the Minecraft?"—"No, here. In the code! Can you see it?"—"No!" Referencing objects and regions on a shared screen during remote collaborations can be frustrating. Some teams, however, perform surprisingly well under the constraints of the virtual environment. Here, we test what gaze-UI coupling and regularity–the alignment and collective patterns of teammates attentional focus in conjunction with changes in a user interface–reveal about the collaboration process and outcomes.

Effective collaboration relies on communication skills, which underlie and support mutual sharing of concepts, checking assumptions, calibrating intentions, and monitoring progress. Collaboration in computer-mediated settings, however, changes communication processes and generated and transmitted signals compared to face-to-face scenarios [50]. Thus, collaborators have to alter their style of communication to compensate for the lower quality social and behavioral signals. Here, we study patterns of visual attention when small groups engage in collaborative problem solving in a computer-mediated context and their relationship to quality of collaboration and its outcomes.

Motivation

Good interpersonal and communication skills are crucial in digital environments, particularly those that require collaborative problem solving [1]. Groups can outperform an individual in problem-solving tasks, but more often than not demonstrate "process loss" by failing to live up to their potential [21, 26, 51]. One type of process loss—coordination loss [18, 36, 58]—occurs when key socio-affective-cognitive (SAC) processes, such as conversation grounding, turn taking, emotion co-regulation, and joint attention, which contribute to overall collaboration quality [3, 15, 20, 26, 27, 32] are disrupted. Further, SAC processes are often muted during remote collaborations [2, 48], and current computer-supported cooperative work (CSCW) or cooperative learning (CSCL) environments rarely provide supports for them.

In terms of visual attention, which is the focus of this work, previous research demonstrates that coupling of an individual's attention with others is indicative of how he or she follows the conversation, interaction, and their intentions (e.g., [50]). However, most of the work has been done in the context of dyads engaged in a conversation [23, 42, 46]. Furthermore, a significant portion of research focuses on the extent to which dyadic eye gaze is aligned when individuals are attending to the same stimuli [23, 42, 43]. However, this work does not recognize

that tasks often require team members to adopt different roles and attend to different features of the task in order to promote performance. As a result, we know little about multiparty visual attention, how it emerges over the course of an interaction, and how it influences the quality of collaboration. Accordingly, we extend prior work on dyadic gaze alignment by investigating collective patterns of gaze in context with the activity on a shared screen during computer-mediated collaborative problem solving.

The landscape drastically changes when small groups are considered. Decades of research on small group interactions has concluded that fundamentally different phenomena emerge when one goes beyond dyads because "dyads are not groups" [33]. Dyads are largely limited with respect to the team dynamics they can afford, whereas complexity rapidly scales up when teams expand. While dyads afford three interacting units (two individuals and one pair), triads involve multiple and more complex interacting units (three individuals, three dyads, and one triad), thus increasing demands on group coordination. In addition, a range of interesting dynamics arise when triadic eye gaze is coupled with the interaction context, as in this paper. In that case, the analysis considers four simultaneous multivariate signals (eye gaze of three team members (A,B,C) and changes to the user interface (I)), resulting in 11 possible interacting components (AB, AC, BC, AI, BI, CI, ABC, ABI, ACI, BCI, ABCI). In this work, we investigate *gaze-UI coupling* (AI, BI, CI) and *gaze-UI regularity* (ABCI).

Research questions and novelty

To understand these complex collaborative interactions, we model visual attention as a nonlinear and interlocked process with spontaneously emerging spatio-temporal patterns situated within the unfolding interaction. Using cross- and multidimensional- recurrence quantification analyses (CRQA and MdRQA, respectively), we investigate whether groups exhibit systematic patterns during collaborative problem solving and whether these patterns predict collaborative processes and outcomes.

Our work is novel by providing a computational assessment of complex computer-mediated collaboration from the lens of triadic visual attention. Our approach is grounded in the theory that teams can self-organize into interpersonal synergies [11, 17], which allows interacting individuals to function as a single unit, reducing the number of degrees of freedom (dimensional compression) and allowing individual components to rapidly react to changes in others (reciprocal compensation) [44].

Our main research question centers on how group-level gaze-UI regularity, defined as the dynamics of teams'

collective attention aligned with changes in the UI, predicts collaborative problem-solving processes and outcomes after accounting for individual-level gaze coupling with the UI and other covariates (e.g. verbosity). Specifically, we expand on previous research investigating visual attention during collaborative tasks in three key ways. We are the first to establish a measure of triadic visual attention in computer-mediated collaborative problem solving. Second, combining two distinct variants of RQA, we quantify spatial-temporal patterns between intra-individual's gaze and UI changes (individual-level gaze-UI coupling via CRQA) and inter-individuals' gaze along with UI changes (group-level gaze-UI regularity via MdRQA). Third, we show that our RQA measures of predict measures of collaboration quality in terms of objective (i.e., task scores and expert-coded measures of effective collaborative problem solving) and subjective assessments of the quality of the collaboration and its outcome.

Our results indicate that not only do teams align eye gaze with the changing environment and produce regular group-level patterns of gaze, the extent to which they do so positively predicts collaboration quality and outcomes. Thus, patterns of visual attention within teams aligned with a UI might be a key instrumental variable to consider in next-generation intelligent user interfaces that aim to improve collaborative outcomes by monitoring the unfolding collaborative process and dynamically intervening as needed.

2 BACKGROUND

Visual attention in collaborative interaction

Research examining eye gaze during collaborative tasks often focuses on joint attention, or the alignment between partners' locus of attention. Joint attention underlies successful communication [3] and is crucial for active engagement and participation in conversation [24, 41]. Conversely, the lack of joint attention has been linked to misunderstandings [7]. Indeed, attending to the same areas of interest promotes collaborative flow and mutual understanding [26] and has been associated with improved problem solving and higher learning gains [40].

To foster joint attention during computer-mediated tasks, researchers have experimented with gaze visualizations and real-time sharing of gaze between two remote partners in a number of domains, including remote pair programming [5, 13, 24], collaboratively solving puzzles [14], collaborative learning of neuroscience concepts [46], and visual search tasks [6, 35, 60]. In these studies, the gaze of one team member is shown to their partner in near-real time. Gaze sharing allows for implicit

referencing without the need of verbal cues, improving awareness of partner's locus of interest [60], action coordination [13], and mutual understanding of stimuli [46].

Thus far, research on visual attention has mainly been restricted to studies on dyadic gaze synchrony and on how joint visual attention is associated with performance [23, 42, 43]. It is unclear how findings will generalize to small groups of three or more and whether going beyond gaze alignment to collective patterns of gaze will provide new information about the collaborative process and outcomes. Here, we adopt a dynamical systems framework and associated analytic techniques to understand visual attention during triadic problem solving in a computer-mediated setting.

Dynamical systems to analyze human interactions

Dynamical systems frameworks view human interaction as a continuous and mutually adaptive process [8, 19, 29, 45]. Synergies are one type of nonlinear dynamical system that occur when interacting components function as a single unit. A synergy arises as the system's degrees of freedom become loosely coupled and mutually constrain each other. Thereby, the shared set of possibilities gets temporarily reduced and allows for more stable and *regular* behavior [44, 50]. In addition, synergies adaptively re-organize to create novel or *irregular* behaviors in order to meet shifting task demands.

From the perspective of synergies, effective collaboration is supported by team members' ability to coordinate and coregulate one another. Coordination refers to the functional organization of interacting parts and processes that give rise to spatial and temporal order [25]. Relatedly, coregulation refers to the process by which responses of one component change based on those of another component. Coordination and coregulation give rise to emergent patterns of behavior that are not reducible to the activity of its individual components, such that the whole is greater than the sum of its parts. In the context of teamwork, coordination and coregulation may occur across multiple interacting channels in order to promote task performance [38, 53]. Here, we focus on one interacting channel-pattern of eye gaze as an index into visual attention, and employ an analytic method from dynamical systems theory, called RQA, to uncover synergistic patterns at the team-level.

Recurrence in joint attention

Recurrence Quantification Analysis (RQA) is a nonlinear time series analytic technique for studying dynamic patterns over time [30] and can be used to capture activity common to synergies, including adaptive shifts between regular and irregular patterns of behavior. Contrary to

traditional linear approaches to time series analysis, RQA relaxes requirements of signal linearity and stationarity, which are limits of many (but not all) traditional time series analysis techniques [49]. Cross-RQA (CRQA) is one type of RQA that quantifies coupling between dyadic signals, or the extent to which two signals converge on the same state across time. Multidimensional-RQA (MdRQA) is another and more recent extension of RQA [56, 57] that can accommodate numerous signals which may be from different individuals and/or modalities. Unlike CRQA, MdRQA captures the extent to which collective organization of the system exhibits regular or irregular patterns of behavior, where regularity refers to periods during which the multidimensional system revisits a particular state, though the individual signals may not match one another. In the context of team eye gaze, a stable state may correspond to Participant A monitoring changes in the code while Participant B focusing on another portion of the screen where they view the animation that results from the code. Thus, CRQA can be considered a measure of coupling between two time series, while MdRQA is a measure of system-level regularity.

While MdRQA is a novel analytic tool, CRQA has been used to investigate joint attention, or mutual gaze coupling, between speaker and listener [43], pair programmers [23], and medical experts and novices [55]. CRQA measures have been shown to be indicative of various interaction outcomes. Gaze coupling between programmers, for example, has been associated with better division of labor, mutual participation, and strategy agreement [23]. This study also found that gaze coupling was highest during moments of verbal references, suggesting that verbal interaction was associated with a focusing of eye gaze toward particular aspects of the task. Furthermore, [42, 43] found that coupling between speaker and the listener was predictive of listener's comprehension.

Despite showing promise in predicting and assessing various aspects of collaboration quality, prior research utilizing CRQA has been limited to analyzing the joint attention between dyads. However, measures of joint attention neglect the fact that team members may need to focus on different features of the task environment. This may be evident in systematic group-level patterns of eye gaze that may not include attending to the same area of interest at the same time. To our knowledge, we present the first study of the dynamics of team's visual attention aligned with changes in the user interface (UI) and its influence on collaboration quality.

3 DATA SOURCE: HOUR OF CODE

We used data collected as part of a larger project on collaborative problem solving [52]. In this work, teams of students used a video-conferencing tool to collaboratively solve a challenging task in a Minecraft themed Hour of Code[1]. The interface uses a visual programming language where each code chunk is represented as a syntactically correct interlocking blocks available in the block library (Figure 1b). The participants were tasked with building a brick building in the Minecraft world (Figure 1a) using the code blocks (Figure 1c).

Participants

Participants were 111 undergraduates from a medium-sized private Midwestern university. Participants were 74.8% Caucasian, 9.9% Hispanic/Latino, 8.1% Asian, 0.9% Black, 0.9% American Indian/Native Alaskan, 2.7% other, and 2.7% did not report ethnicity. The average age was 19.4 years and 63.1% of the participants were female. Participants were compensated with course credit.

Participants were assigned to 37 teams of three. The only inclusion criterion was no previous experience with computer programming and none of the participants were excluded on this basis.

Experimental task and procedure

Participants were randomly assigned to one of three computer-enabled rooms in a lab. Each computer was equipped with a webcam and microphone for video conferencing with screen-sharing through Zoom[2]. Participants' eye movements were recorded with the EyeTribe eye-tracker, with binocular sampling at 30Hz, mounted on a tripod below the screen. The screen content was also recorded using Zoom's built in features.

Upon arrival in the lab, participants completed individual surveys gathering basic demographic data as well as self-reported standardized test score (ACT/SAT), which strongly correlate with actual test scores [10]. Then, participants were randomly assigned to their task roles. One team member controlled the group's interaction with the environment. The other two team members viewed the controller's screen via screen sharing and actively provided suggestions throughout the collaboration. In order to learn basic programming principles and familiarize them with the Minecraft environment, teams completed five lessons and watched three accompanying videos within the first 20 minutes.

After completing the introductory tasks, participants individually (i.e., screen sharing was turned off) rated their level of satisfaction with their team's "performance

---

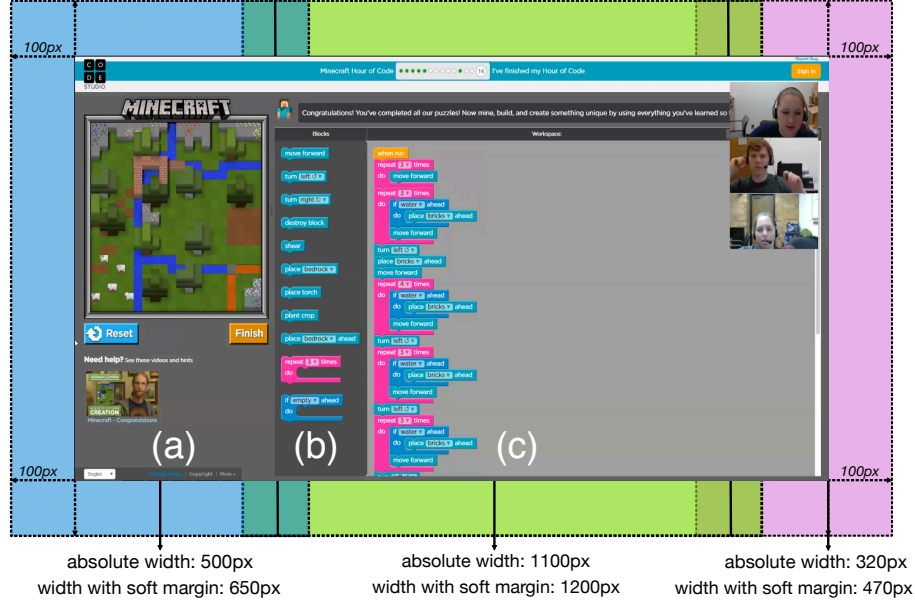[1]https://code.org/minecraft
[2]https://zoom.us

Figure 1: Virtual environment of Minecraft Hour of Code. Areas of Interest (AOI) (with overlap) refer to areas of Minecraft animation (blue), Code (green), and teammates (magenta).

at completing the lessons", how well their team "communicated with each other", how well their team "cooperated to complete the lessons", and how "agreeable my teammates are" using a scale ranging from very dissatisfied (1) to very satisfied (6).

Teams then collaborated on the main task where they had 20 minutes to build a four by four brick building with the following constraints: use at least one if statement; use at least one repeat loop; build at least three bricks over water; and use 15 blocks of code or less.

After completing this challenging task, participants individually rated their subjective assessments of their team's performance, communication, cooperation, and agreeableness using the same items and scales as above. Then, participants individually completed a ten-item multiple-choice test to assess their conceptual knowledge of coding concepts. Finally, participants rated the extent to which they knew each of their teammates (prior to the study) on a scale of 1 (not at all) to 5 (very well).

Measures

We analyzed a number of personal, interpersonal, subjective, and objective aspects of the collaborative process and its outcomes. The following measures of interest were obtained for the challenging task and not for the introductory tasks.

Expert-coded measures. We obtained expert codings of each participant's collaborative problem-solving competencies characterized by construction of shared knowledge, negotiation and coordination, and maintaining team function. These three facets were derived from a theoretically-grounded and empirically-validated model of collaborative problem-solving competencies [53].

Construction of shared knowledge is defined as team members expressing their ideas and working to understand others' ideas. Negotiation and coordination pertains to teams discussing and arriving at a solution and then executing it. Maintaining team function is related to sustaining productive dynamics. Each of these facets has subfacets which in turn have a set of behavioral indicators. For example, constructing shared knowledge is composed of two subfacets: 1) shares understanding of problems and solutions and 2) establishes common ground. Proposing a specific solution is one positive indicator of the first subfacet, whereas a team member interrupting or talking over others is a negative indicator of the second subfacet.

We used a thin slicing approach from [34, 37]. Specifically, each 20-minute video of the challenging programming task was split into four five-minute windows. Within each five-minute window, we selected a random 90-second interval for coding, where a coder counted the occurrence frequency of each indicator per participant. Thus, 360 seconds (30%) of each video was coded.

Through iterative processes of coding and discussion, three coders reached an agreement of 0.92 (Spearman correlation) on a random sample of three videos. The reliability was obtained by first computing indicator-level reliability across the three pairs of coders and then averaging across indicators. Then, the three coders divided the videos among themselves and individually coded their assigned videos.

We aggregated the indicator-level codes across the four 90s segments to obtain participant-level facet scores. We first z-scored the frequency counts by coder to account for variation among the coders. The z-scores were averaged across segments within each participant for a participant-level score. Then, we z-scored each indicator across all participants in order combine indicators to compute sub-facet scores. We multiplied negative indicators by -1, averaged indicator scores within each subfacet, and then averaged subfacet scores to the appropriate facet.

Objective Measures. Two independent raters scored each team's solution on the five challenge criteria, with each criterion being worth one point. Task scores could range from 0 to 5 (M = 2.86, SD = 1.06). The two raters reconciled any disagreements via discussion. We also averaged the 10 individual post-test items to obtain a measure for individual learning (M = 43%, SD = 15.7).

Subjective Measures. Participants' self-reports of their teams' performance was taken as their subjective measure of collaboration outcomes. We averaged each participant's self-reports of communication, cooperation, and agreeableness because ratings were highly correlated (Cronbach's alpha = .89), thereby obtaining a single measure of their subjective perceptions of their team's collaboration process.

Verbosity. We used the IBM Watson Speech to Text service[3] to generate transcriptions of each participant's audio recordings. We summed all the words in the transcript to obtain a participant-level measure of verbosity.

Data preprocessing

We focus on eye gaze behavior in relation to interaction on the screen. We processed the eye gaze data as follows. First, the raw gaze points were categorically assigned to one of the following four areas of interest (AOIs): Minecraft animation, Code, Partners and Out of bounds (Figure 1). We chose to work with raw gaze points rather than fixations as this resulted in lower data loss. To compensate for potential eye-tracker calibration errors, we enlarged AOIs by a 100px safety margin around the screen and by a 50px margin between the

---

[3]https://www.ibm.com/watson/services/speech-to-text

AOIs, as illustrated in Figure 1. Gaze points in the safety margin were assigned to both areas. Figure 2 illustrates gaze distribution to the designed areas, where we note that majority of eye gaze was on the Code and then on the Minecraft area. The area with the live view of teammates was only sporadically visited, probably because of its small size. Interestingly, proportions of eye gaze towards the AOIs were quantitatively similarly distributed across roles. Next, the stream of assigned AOIs was segmented into 1-second time windows and each segment was assigned to the majority AOI.
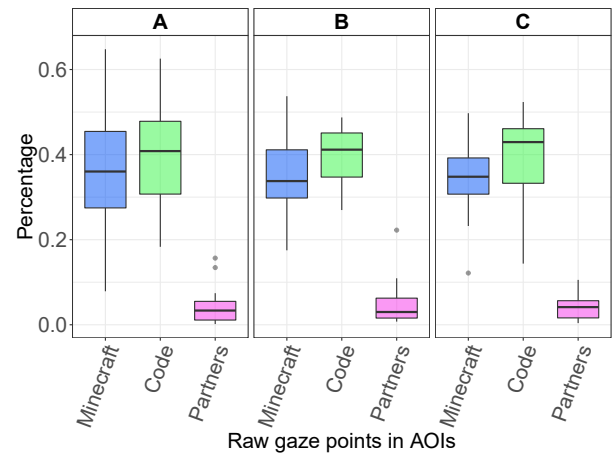

Figure 2: Distribution of raw gaze points across participants and their roles. Participant A controlled the environment.

Because the task was hosted by a third party website, activity log-files were not available. Therefore, we processed screen videos to obtain a measure of activity on the screen. Specifically, we used a validated motion estimation algorithm [58] to compute the proportion of pixels that changed for each AOI and aggregated the data in one-second time windows. Thus, each team was represented by three (one for each team member) time series of gaze AOIs (Minecraft, Code, Partners, Out-of-bounds) and one time series of screen AOI changes (Minecraft, Code, Partners).

Finally, we removed AOI data points associated with missing data caused by the eye tracker failing to register gaze to ensure that RQAs were grounded in the same data sources. We decided to only include participants with at least 75% non-missing raw data (57 participants across 19 complete triads). Data loss was expected due to the low-cost sensor with no chin rest, which was important for ecological validity. We aimed to avoid spurious results due to missing data; therefore, whenever one of the partner's time series was missing an AOI value, the corresponding AOI values of the other two partners were

removed. Across teams, the rate of cleaned sequences ranged from 8.9% to 66.5% (M = 41.9%, SD = 16.8). We included missing data proportions as covariates in the models and this had no statistically significant influence on reported results.

4 RECURRENCE QUANTIFICATION ANALYSIS

Recurrence Quantification Analyses unveil the temporal organization of time series by identifying recurrent patterns, or repeated values, that occur over time [30]. Auto-Recurrence Quantification Analysis (ARQA) is limited to examining patterned behavior of a single signal (i.e. time series of a single participant), Cross-Recurrence Quantification Analysis (CRQA) identifies patterns of coupling between two signals (e.g., a participant and activity on the screen in our case), and Multidimensional Recurrence Quantification Analysis (MdRQA) can reveal collective patterns of regularity across multiple data streams—in our case, three participants' AOIs and screen activity AOIs.

CRQA and MdRQA in principle

CRQA can identify the degree to which two time-series trajectories are in a similar state for one time point at a particular time lag, as well as uncover longer repetitive patterns across a series of co-occurring sequential points. In this way, various information about interaction patterns can be quantified by determining how often two time series co-visit one another [9].

We use CRQA to examine coupling between each participant's eye gaze and changes in screen activity. This serves as a baseline coupling measure since it is independent of the other teammate's eye gaze. In addition, we utilize MdRQA, an extension of RQA for multiple time series (i.e., > 2) [57], to assess repeated patterns of team-members' visual attention in context with screen activity (4 time series). Unlike CRQA where recurrent points correspond to matching values between two time series, recurrent points in MdRQA represent collective states across time series where the individual time series might not be in alignment with one another but the overall configuration of the system is repeated (e.g., a stable configuration might include system A being "on" and system B being "off"). Thus, we utilize MdRQA to measure group-level gaze regularity.

For CRQA, the two time series are first transformed into a distance matrix representing the pairwise Euclidean distances between elements of each series. The axes of the resulting distance matrix correspond to the two time-series, such that each cell in the distance matrix is the distance between time points from time series x and y at a particular time lag. For MdRQA, the distance matrix

reflects the Euclidean distances between elements within each time series at different time lags. In other words, the data is treated as a multidimensional times and the matrix represent distance between multidimensional time series elements at various time lags [57]. Figure 3 illustrates an example how a color-coded MdRQA recurrence plot is created from gaze-UI time series.

In both CRQA and MdRQA, the diagonal of the matrix is the line of identity (LOI), or the distance between elements of time series that occur at the same time point. Diagonal lines parallel to the LOI represent points that occur at a particular time lag, with lines further from the LOI representing greater time lags between x and y time points. Because MdRQA compares each time series to itself across various lags , the LOI (lag 0) is always 0 and is excluded from analyses.

Finally, for both CRQA and MdRQA, a radius is applied in order to transform the distance matrix into a recurrence matrix. The radius defines whether two points are sufficiently similar to one another enough to constitute a "match". If the distance in a given cell of the distance matrix is less than the radius, the cell is identified as a recurrent point and is assigned a value of 1 in the recurrence matrix. Distances between x and y time points that are equal to or exceed the radius are assigned a value of 0.

CRQA and MdRQA measures

A variety of measures are used to quantify patterns in a recurrence matrix [42, 56]. Since the measures are highly correlated, we focus on two here. *Recurrence rate* (RR) is the percentage of recurrent points in the recurrence matrix points [40, 54]. We use CRQA recurrence rate to index the degree of coupling between individual participants' eye gaze and the UI activity, and we use MdRQA recurrence rate to index group-level gaze regularity among triads and the UI.

Entropy of diagonal line lengths (ENTR) is calculated based on the distribution of diagonal lines on the recurrence matrix where two or more sequential diagonal points are considered to form a "line." These lines correspond to periods of time where the system repeats sequences of activity and longer lines correspond to longer sequences. Entropy is an index of complexity where deterministic systems yield less complex dynamics and have lower levels of entropy, such that the trajectory of a system is more predictable [39]. On the other hand, less deterministic systems have greater entropy, as they are more disordered and exhibit greater complexity [28].

RQA parameters and analysis

Both CRQA and MdRQA were set with default parameters for categorical input [9, 49]: delay = 1, radius= 0.0001,
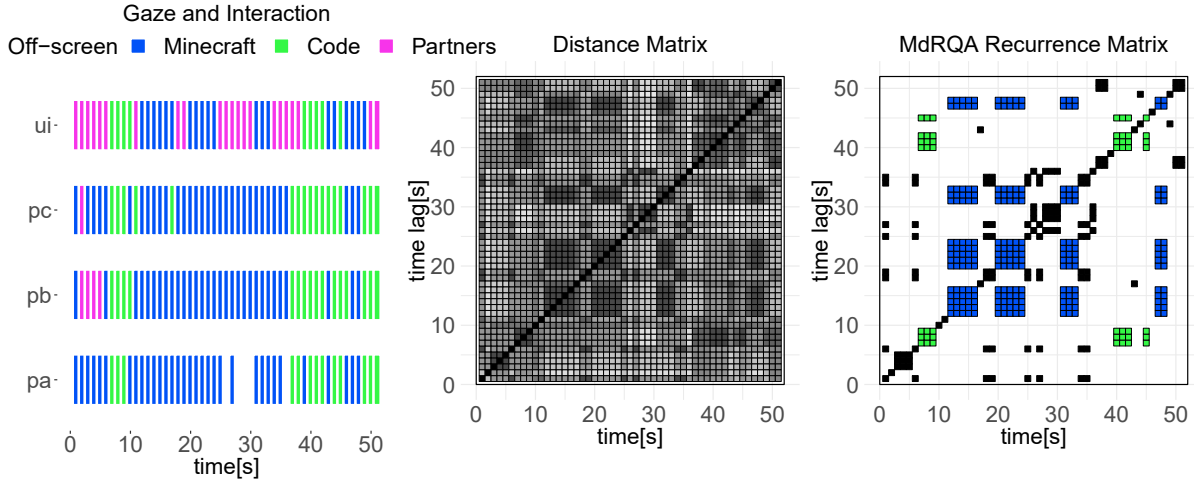
Figure 3: Construction of MdRQA recurrence matrix. Time series of UI changes and gaze of participants A, B, and C are assigned to Areas of Interest (Minecraft, Code, Partners). The distance matrix is computed using the 4-dimensional Euclidean distance between the time series elements across different time points. For example, cell [1, 5] corresponds to the distance between gaze-UI elements at time points 1 and 5. A radius is applied to the distance matrix to produce a recurrence plot. The colored points illustrate when the focus of all participants is both recurrent and aligned in the same AOI, either for the Code (green) or Minecraft animation (blue). Points in black illustrate to other AOI combinations is recurrence but not aligned gaze.

embedding = 1, norm = 0 (Euclidean distance). Data pre-processing was performed in Python using Pandas [31]. The RQA measures were computed using R packages 'crqa' [9] and 'mdrqa' [56]. Further analyses were conducted using R packages 'lme4' [4], 'car' [16], and 'ggplot2' [59].

5 RESULTS

We examined the predictive value of gaze synchrony on the three expert-coded collaboration quality measures (construction of shared knowledge, negotiation coordination, and maintaining team function), subjective (perceptions of team performance and collaboration), and objective (post-test scores and task score) outcomes.

Gaze-UI coupling and regularity versus baseline

We first compared observed CRQA measures to the average of five baselines. Each baseline was created by randomly shuffling the observed gaze time series while keeping the series of screen activity unchanged, before submitting the shuffled time series to CRQA. The average recurrence rate and entropy across the repetitions was used as baselines [56]. This baseline method is beneficial because shuffling maintains the distributional characteristics of the original time series while breaking the temporal dependencies [12].

Overall observed vs. chance recurrence are expected to be identical since every recurrent point will reoccur with every other point, just not systematically. Although the total number of recurrent points in the plot will be the same, the points should not be systematically patterned in the shuffled baseline, which the entropy measure should detect. Accordingly, we conducted paired-samples t-tests to compare observed vs. baseline entropy and found higher entropy of diagonal line lengths for the observed ($M = 1.47$, $SD = 0.20$) vs. the baseline ($M = 1.00$, $SD = 0.17$) time series, $t(72) = 23.55$, $p < .001$.

Baseline MdRQA was computed from gaze AOIs of all three participants and the screen activity AOIs so that concurrent values remained together but the temporal order was shuffled. Similarly to the CRQA baseline, overall recurrence rate of the baseline MdRQA will be same to the observed MdRQA. However, entropy of diagonal line lengths of observed behavior ($M = 0.95$, $SD = 0.19$) were significantly higher than the shuffled baseline ($M = 0.32$, $SD = 0.07$); $t(18) = 17.80$; $p < 0.001$.

Figure 4 illustrates diagonal recurrence rates (i.e., the percentage of recurrence points on a particular diagonal) from MdRQA recurrence matrices in comparison to shuffled baselines for 10 lags (i.e., 10 lines parallel to the diagonal). While baseline diagonal recurrence rates fluctuate between 0.05 and 0.1, observed diagonal recurrences shows a systematic pattern that are lowest at the farthest lag of 10 seconds, but gradually increase for smaller lags, until they peak around the line of identity. Thus, team gaze in the context of screen activity exhibits greater regularity during relatively short 1-2 second time lags.
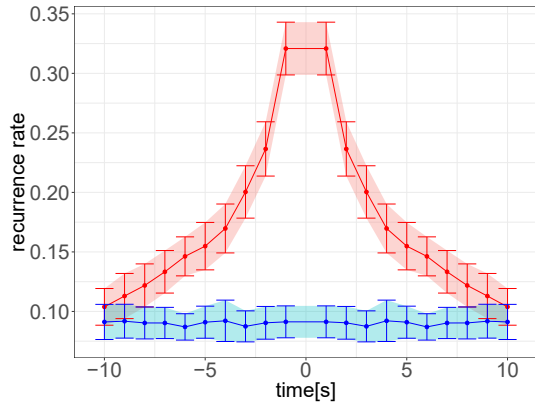
Figure 4: MdRQA diagonal recurrence rate with positive time lags in observed gaze-UI regularity (red) against randomized baseline (blue). Error bars illustrate 95% confidence intervals.

Gaze-UI coupling and regularity as a predictor of collaboration quality and outcomes

We examined the predictive value of gaze-UI coupling and regularity on the three expert-coded collaboration quality measures (construction of shared knowledge, negotiation/coordination, and maintaining team function) and on subjective (perceptions of team performance and collaboration) and objective (post-test scores and task score) outcomes.

Participant-level analyses. We used linear mixed-effect models to predict each of the outcome variables (except for task scores, which is a team level variable) with team as a random intercept, a participant-level CRQA measure as a level 1 (participant-level predictor), and a team-level MdRQA measure as a level 2 (group-level) predictor. This is the recommended analytic technique due to inherent nesting in our design where participants are nested within teams [4]. We included recurrence rate and diagonal line length entropy in separate models. However, we included CRQA and MdRQA versions of each in the same model to ascertain the incremental predictive validity of team-level regularity (e.g., MdRQA recurrence rate) after accounting for participant-level coupling (e.g., CRQA recurrence rate).

We also included a number of covariates to account for confounding factors. These were: whether the participant was the one who controlled the interface (1 or 0), ACT/SAT scores(to control for scholastic aptitude), whether the participant was familiar with his or her other team members (1 or 0), total words spoken by the participant (to control for verbosity), and the percentage

of valid data (to control for biases associated with eye-tracking quality).

We found that after controlling for covariates (see Tables 1 and 2), CRQA recurrence rate significantly predicted student's construction of shared knowledge skills ($B = 0.050$, $p = 0.011$), and MdRQA recurrence rate predicted significantly student's negotiation and coordination skills ($B = 0.042$, $p = 0.016$). Entropy of diagonal line lengths from CRQA was also predictive of negotiation and coordination ($B = 0.509$, $p = .065$).

Whereas neither CRQA nor MdRQA recurrence rate predicted the subjective and objective outcomes, CRQA entropy was a negative predictor of subjective evaluations of team collaboration ($B = -1.742$, $p = 0.053$). Importantly, both CRQA entropy ($B = -2.054$, $p = 0.035$) and MdRQA entropy ($B = 1.991$, $p = 0.082$) predicted subjective perceptions of team performance, but in opposite directions.

In addition to gaze-UI regularity, we were interested in UI-team-synchrony, which we computed based on whether eye gaze of all three teammates aligned and also corresponded to the active UI AOI (gaze-UI synchrony) vs. when eye gaze of all teammates aligned but did not correspond to the active UI AOI (gaze-only synchrony). We filtered the MdRQA matrices based on these two criteria and computed RQA metrics on the filtered matrices. We found that MdRQA recurrence rate was positively correlated with both gaze-UI synchrony (r = 0.650, p= 0.003) as well with gaze-only synchrony ($r = 0.748$, $p < 0.001$). Thus, between 43% to 55% of the variance in gaze-UI regularity can be accounted by gaze-UI and gaze-only, synchrony respectively.

Next, we re-ran the regression models with recurrence rates corresponding to the two synchrony measures in lieu of gaze-UI regularity. We found that both measures positively predicted negotiation and coordination similarly to regularity (recurrence rate of gaze-UI synchrony: $B = 0.050$, $p = 0.084$; recurrence rate of gaze-only synchrony: $B = 0.042$, $p = 0.050$), essentially replicating the finding with regularity. In addition, recurrence rate of gaze-only synchrony predicted perceptions of performance ($B = 0.152$, $p = 0.077$). Gaze-UI and gaze-only synchrony-based entropy of diagonal line lengths did not predict any of outcomes.

Team-level outcome analyses (Task score). We also examined the relationship between team-level gaze-UI regularity and team outcome measures-namely task score. For this analysis, we averaged each individual's CRQA measures per triad to obtain team-level averages. MdRQA measures and task score were assessed per team and are

Table 1: Subject-level LME: Recurrence rate

	Dependent variable:					
	Post-test	Perception of team performance	Perception of team collaboration	Constructing shared knowledge	Negotiation Coordination	Maintaining team function
CRQA RR	0.007 (0.055)	−0.050 (0.051)	0.041 (0.045)	0.050** (0.019)	0.003 (0.014)	−0.003 (0.021)
MdRQA RR	−0.020 (0.073)	0.088 (0.072)	0.007 (0.074)	0.005 (0.024)	0.042** (0.017)	−0.006 (0.025)
Covariates:						
ACT	−0.119 (0.171)	−0.194 (0.144)	−0.084 (0.118)	−0.057 (0.069)	0.025 (0.047)	0.043 (0.075)
Controller	−0.212 (0.276)	0.350 (0.228)	0.033 (0.184)	0.361*** (0.115)	−0.065 (0.078)	0.103 (0.128)
Familiarity	−0.169 (0.492)	0.009 (0.493)	−0.153 (0.515)	0.114 (0.157)	0.032 (0.115)	−0.178 (0.165)
Data validity	0.950 (1.105)	0.722 (1.112)	−0.329 (1.168)	−0.205 (0.348)	0.326 (0.256)	0.331 (0.366)
Verbosity	0.289* (0.158)	−0.219* (0.133)	0.123 (0.108)	0.191*** (0.064)	0.107** (0.044)	0.115 (0.070)
Constant	−0.220 (1.351)	−0.254 (1.296)	−0.951 (1.266)	−1.269*** (0.450)	−0.574* (0.326)	−0.012 (0.478)

Note: n=57 $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2: Subject-level LME: Entropy of diagonal line lengths

	Dependent variable:					
	Post-test	Perception of team performance	Perception of team collaboration	Constructing shared knowledge	Negotiation Coordination	Maintaining team function
CRQA ENTR	0.417 (1.041)	−2.045** (0.971)	−1.742* (0.903)	0.257 (0.373)	0.509* (0.276)	−0.251 (0.367)
MdRQA ENTR	−0.192 (1.175)	1.991* (1.146)	1.217 (1.114)	0.238 (0.402)	−0.024 (0.309)	0.034 (0.391)
Covariates:						
ACT	−0.104 (0.167)	−0.248* (0.135)	−0.068 (0.113)	−0.028 (0.071)	0.025 (0.045)	0.036 (0.073)
Controller	−0.208 (0.274)	0.341 (0.219)	−0.007 (0.181)	0.343*** (0.121)	−0.067 (0.075)	0.105 (0.127)
Familiarity	−0.158 (0.490)	−0.131 (0.484)	−0.094 (0.478)	0.119 (0.166)	−0.041 (0.129)	−0.158 (0.161)
Data validity	0.972 (1.082)	0.526 (1.062)	−0.216 (1.044)	−0.365 (0.370)	0.020 (0.285)	0.416 (0.361)
Verbosity	0.281* (0.148)	−0.180 (0.119)	0.085 (0.099)	0.156** (0.064)	0.117*** (0.040)	0.111* (0.066)
Constant	−0.680 (1.290)	0.538 (1.248)	1.354 (1.206)	−0.614 (0.446)	−0.703** (0.340)	0.156 (0.436)

Note:n=57 $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

reported as original values. We used a Pearson correlation rather than a multiple regression framework due to the small number of teams ($n = 19$). Our main finding was that MdRQA entropy of diagonal line lengths was positively correlated with task scores ($r = 0.425$, $p = .069$). Although the correlation was marginally significant at the .05 level, this can be attributed to the small sample size; a correlation of 0.456 would be significant with our sample [22]. CRQA entropy ($r = 0.025$, $p = 0.918$), CRQA recurrence rate ($r = -0.055$, $p = 0.822$), and MdRQA recurrence rate ($r = 0.125$, $p = 0.608$) were not correlated with task score.

Visualizing recurrence plots.

Lastly, we demonstrate how MdRQA recurrence plots can be used to provide a qualitative overview of teams' gaze-UI dynamics. Enhanced recurrence plots illustrate not only the overall distribution of recurrent points but also color-coded recurrent points associated with gaze-UI synchrony on a particular AOI. Red and yellow regions illustrate moments when all participants were fully and partially (triad-only) aligned with the AOI, respectively, while the black regions highlight moments when a team exhibited other collective gaze patterns-UI patterns. For a comparison, we selected two teams, which scored high (score = 0.498) vs. low (score = -0.314) on the experts' assessments of negotiation and coordination. The team, which effectively negotiated solutions and coordinated their execution (5a), also exhibited higher gaze-UI regularity and full gaze-UI synchrony in the Code area. The team with the lower skill of negotiation/coordination exhibited a lower number of recurrent points and also fewer points associated with full synchrony, mainly in the area of the Minecraft animation.

(a) High negotiation and coordination

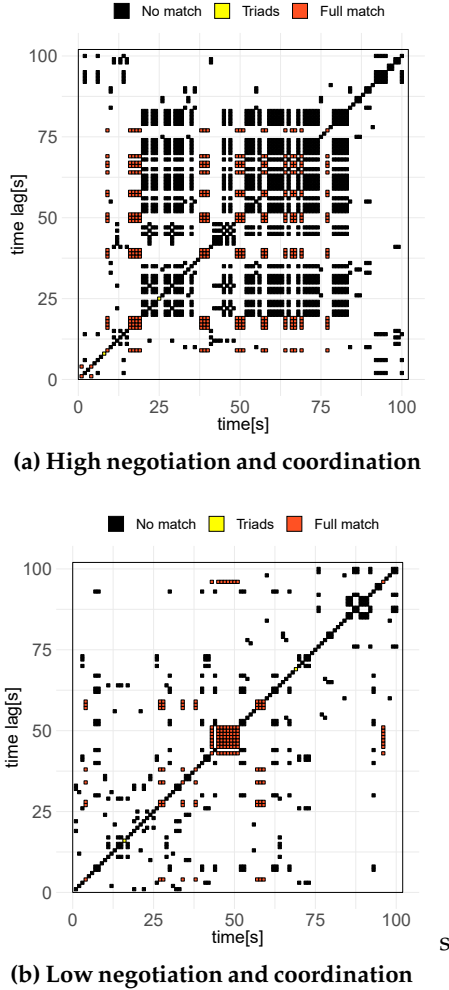**(b) Low negotiation and coordination**

Figure 5: Example multidimensional recurrence plots of two teams, which scored high and low in negotiation and coordination. Red points represent moments of gaze-UI synchrony when all three participant's gaze and activity in the UI align, and yellow points represent gaze synchrony when gaze of all three participants is aligned but different from the active UI region. Black points indicate regular recurrent patterns without full synchrony.

6 DISCUSSION

We investigated systematicity in patterns of triadic visual attention towards ongoing interaction on a shared screen during collaborative problem solving and assessed whether these patterns were predictive of collaboration quality, subjective outcomes, individual learning and team's task performance. Using a dynamic systems framework and associated analytic tools of recurrence quantification analysis, we quantified not only how a team member individually coupled his or her attention with UI changes

but also repeated patterns in the triad's attention in context with UI changes. We confirmed that participant-level gaze was significantly coupled with screen activity changes, suggesting they were attending to the task. Importantly, team-level recurrence measures significantly differed from randomly shuffled baselines, providing evidence of systematicity in team-attention dynamics. We also found that a bit over half the variance in team-UI gaze regularity was explained by gaze-synchrony, indicating that there are patterns in visual attention dynamics that go beyond mere synchrony, which has been the focus of much previous work [23, 46, 47]. Taken together, this suggests that a teammate's eye gaze exhibit significant patterns of coupling with the UI as well as other teammates along with and the triadic gaze-UI system shows patterns of regularity.

But does the degree of coupling and regularity matter or is it merely incidental? Previous research on dyadic interaction has presented correlational evidence on joint attention and interpersonal coordination during dyadic interactions. Our results go beyond findings on dyads and provide more detailed insights on multiparty collaboration. We found that team-level patterns of gaze-UI regularity underlie different collaboration qualities and outcomes beyond participant-level gaze-UI coupling. Specifically, we found that the coupling (via recurrence rates) between participant's eye gaze and the ongoing screen activity was a positive predictor of constructing shared knowledge. This is an intuitive finding because attending to screen activity is necessary for providing suggestions that contribute to a problem solution. In contrast, gaze-UI regularity at the team level was a positive predictor of negotiation and coordination, suggesting that teams who exhibited more consistent patterns in gaze-UI dynamics were more effective on this component. Gaze-only synchrony, a component of gaze-UI regularity, also predicted negotiation/coordination suggesting that to effectively negotiate solutions and coordinate execution, teams must be on the same page in terms of their momentary attention with each other and should show some consistency in gaze dynamics. This measure also predicted perceptions of task performance, confirming prior dyadic studies on joint attention.

Conversely, increased entropy of participants' gaze-UI coupling was associated with elevated negotiation and coordination skills. In this context, diagonal lines correspond to consecutive gaze-UI couplings and their lengths describes duration of the coupling sequence (e.g., "Code-Code → Minecraft-Minecraft → Code-Code" has length of 3 seconds). Increased gaze-UI coupling via entropy of diagonal line lengths suggests that the elevated negotiation and coordination promoted coupling

sequences of diverse durations. Interestingly, this measure was negatively associated with his or her perceptions of team performance and of the collaboration. It is possible that frequent and diverse gaze-UI coupling sequences caused participants discomfort, and, in turn, concluded that their collaboration was less smooth, and the performance of the team was subpar.

At the team-level, conversely, entropy of gaze-UI regularity positively predicted perceptions of team performance. This is not necessarily a contradicting finding since team-level gaze-UI regularity does not require an exact match of participants' attention to the UI unlike participant-level gaze coupling. Instead it indexes more variability in the lengths of recurrent gaze-UI sequences, which is an indicator of novel behavior. This measure also correlated with task-scores, further confirming the benefits of novelty.

Finally, neither participant-level gaze-UI coupling nor team-level gaze-UI regularity were significantly predictive of individuals' learning or teams' task scores. Thus, despite visual attention being a key component in collaboration, it only partially revealed complex cognitive processes in collaborative problem solving, at least with respect to the measures considered here.

Applications

The present findings suggest that measures of teams' visual attention can provide important information about the collaborative process, which may complement objective task scores. These measures may illustrate the "why" and "how" of team interaction since they provide indices of attention and coordination, which index multiple process and outcome measures. In addition to the quantitative findings encoded in recurrence measures, color-coded recurrence plots offer global insights on collaboration during team tasks, which can be used for qualitative assessments of the interaction.

From the perspective of next-generation collaborative user interfaces, dynamics of visual attention in multiparty collaboration present pervasive low-level indicators, suitable for intelligent interventions. Real-time interventions, such as short-term gaze alignment, are particularly critical during remote collaborations with impoverished social signals. Gaze awareness together with coupling and regularity measures can be leveraged to reinforce team efforts and improve outcomes.

Limitations & Future Work

Like all studies, ours has limitations. First, the use of low-cost eye-trackers together with unrestricted range of participants' body and head movements introduced errors in the recorded data. We carefully inspected and rejected the most erroneous cases and analyzed the remaining eye gaze with respect to rather large areas of interest to compensate for this limitation. Next, the low-cost sensor often failed to register participants' eye gaze, which resulted in gaps in the eye-tracking time series. Because leaving such gaps in the eye gaze time series could form spurious recurrent points and artificially increase recurrence rates, we removed time points when one or more team members did not have valid data. Although we controlled for the proportion of the valid data in the models, this is not a substitute for obtaining higher-quality data. Relatedly, although the original data source consisted of 111 participants in 37 teams, eye-tracking data of 19 teams had sufficient quality to be included in our analyses. The reduced number of teams restricted the types of statistical modeling that could be done at the team-level. Thus replication with larger samples and higher quality gaze tracking is warranted, especially in more real-world settings. Finally, we analyzed global measures of gaze-UI-coupling and regularity by computing measures across the entire interaction, In future research, we will investigate other variants of RQAs suitable for real-time assessments and interventions, such as windowed RQAs, which only consider local recurrence by focusing on time lags around the line of identity [47].

7 CONCLUSION

Computer-mediated collaborative problem solving is a highly complex form of interaction where the transmission of certain social, affective, and cognitive states may be impaired in remote settings, reducing mutual understanding between teammates. We investigated triadic computer-mediated collaboration of novice programmers engaged in a collaborative problem solving task. Using measures from a dynamical systems framework, we found that dynamics of visual attention of individual teammates and of the team as a whole assessed in context with UI activity predicted several key variables associated with the collaboration process and its outcomes. The next step is to leverage these findings to develop intelligent collaborative user interfaces that monitor the unfolding collaboration and launching interventions to improve the collaborative process and outcomes.

REFERENCES

[1] 2015. *PISA 2015 Collaborative Problem Solving Framework*. Technical Report. Organisation for Economic Co-operation and Development.

[2] Richard Alterman and Kendall Harsch. 2017. A more reflective form of joint problem solving. *International Journal of Computer-Supported Collaborative Learning* 12, 1 (2017), 9–33.

[3] Sigal G Barsade and Donald E Gibson. 2012. Group affect: Its influence on individual and group outcomes. *Current Directions in Psychological Science* 21, 2 (2012), 119–123.

[4] Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, et al. 2014. lme4: Linear mixed-effects models using Eigen and S4. *R package version* 1, 7 (2014), 1–23.

[5] Roman Bednarik, Andrey Shipilov, and Sami Pietinen. 2011. Bidirectional gaze in remote computer mediated collaboration: Setup and initial results from pair-programming. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 597–600.

[6] 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.

[7] Mauro Cherubini, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2008. Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 173–180.

[8] Jennifer Coates. 1994. 11 No Gap, Lots of Overlap; Turn-taking Patterns in the Talk of Women Friends1. *Researching language and literacy in social context: A reader* (1994), 177.

[9] 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.

[10] James S Cole and Robert M Gonyea. 2010. Accuracy of self-reported SAT and ACT test scores: Implications for research. *Research in Higher Education* 51, 4 (2010), 305–319.

[11] Rick Dale, Riccardo Fusaroli, Nicholas D Duran, and Daniel C Richardson. 2013. The self-organization of human interaction. In *Psychology of learning and motivation*. Vol. 59. Elsevier, 43–95.

[12] Rick Dale, Natasha Z Kirkham, and Daniel C Richardson. 2011. The dynamics of reference and shared visual attention. *Frontiers in psychology* 2 (2011), 355.

[13] Sarah D'Angelo and Andrew Begel. 2017. Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6245–6290. https://doi.org/10.1145/3025453.3025573

[14] Sarah D'Angelo and Darren Gergle. 2016. Gazed and confused: Understanding and designing shared gaze for remote collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2492–2496.

[15] Charles R Evans and Kenneth L Dion. 1991. Group cohesion and performance: A meta-analysis. *Small group research* 22, 2 (1991), 175–186.

[16] John Fox, Michael Friendly, and Sanford Weisberg. 2013. Hypothesis tests for multivariate linear models using the car package. *The R Journal* 5, 1 (2013), 39–52.

[17] Riccardo Fusaroli, Joanna Rączaszek-Leonardi, and Kristian Tylén. 2014. Dialog as interpersonal synergy. *New Ideas in Psychology* 32 (2014), 147–157.

[18] Daniel Gigone and Reid Hastie. 1993. The common knowledge effect: Information sharing and group judgment. *Journal of Personality and social Psychology* 65, 5 (1993), 959.

[19] Jamie C Gorman, Nancy J Cooke, Polemnia G Amazeen, and Shannon Fouse. 2012. Measuring patterns in team interaction sequences using a discrete recurrence approach. *Human Factors* 54, 4 (2012), 503–517.

[20] David A Harrison, Kenneth H Price, and Myrtle P Bell. 1998. Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. *Academy of management journal* 41, 1 (1998), 96–107.

[21] Gayle W Hill. 1982. Group versus individual performance: Are N+ 1 heads better than one? *Psychological bulletin* 91, 3 (1982), 517.

[22] Dennis E Hinkle, William Wiersma, Stephen G Jurs, et al. 1988. Applied statistics for the behavioral sciences. (1988).

[23] 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*. ACM, 1125–1134.

[24] Patrick Jermann and Kshitij Sharma. 2018. Gaze as a Proxy for Cognition and Communication. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 152–154.

[25] Scott JA Kelso. 2009. Coordination dynamics. *Encyclopedia of complexity and systems science* (2009), 1–41.

[26] Norbert L Kerr and R Scott Tindale. 2004. Group performance and decision making. *Annu. Rev. Psychol.* 55 (2004), 623–655.

[27] Heng-Yu Ku, Hung Wei Tseng, and Chatchada Akarasriworn. 2013. Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. *Computers in Human Behavior* 29, 3 (2013), 922–929.

[28] Nikita Kuznetsov, Michael A Riley, and Scott Bonnette. 2013. Nonlinear time series methods for analyzing behavioural sequences. In *Complex Systems in Sport*. Routledge, 111–130.

[29] Joseph H Manson, Gregory A Bryant, Matthew M Gervais, and Michelle A Kline. 2013. Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior* 34, 6 (2013), 419–426.

[30] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. 2007. Recurrence plots for the analysis of complex systems. *Physics reports* 438, 5-6 (2007), 237–329.

[31] Wes McKinney et al. 2010. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, Vol. 445. Austin, TX, 51–56.

[32] Margaret M McManus and Robert M Aiken. 2016. Supporting effective collaboration: Using a rearview mirror to look forward. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 365–377.

[33] Richard L Moreland. 2010. Are Dyads Really Groups? *Small Group Research* 41, 2 (2010), 251–267. https://doi.org/10.1177/1046496409358618

[34] Nora A Murphy. 2005. Using thin slices for behavioral coding. *Journal of Nonverbal Behavior* 29, 4 (2005), 235–246.

[35] Mark B Neider, Xin Chen, Christopher A Dickinson, Susan E Brennan, and Gregory J Zelinsky. 2010. Coordinating spatial referencing using shared gaze. *Psychonomic bulletin & review* 17, 5 (2010), 718–724.

[36] Bernard A Nijstad, Wolfgang Stroebe, and Hein FM Lodewijkx. 2003. Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of experimental social psychology* 39, 6 (2003), 531–548.

[37] Jennifer K Olsen, Vincent Aleven, and Nikol Rummel. 2017. Exploring Dual Eye Tracking as a Tool to Assess Collaboration. In *Innovative Assessment of Collaboration*. Springer, 157–172.

[38] Richard V Palumbo, Marisa E Marraccini, Lisa L Weyandt, Oliver Wilder-Smith, Heather A McGee, Siwei Liu, and Matthew S Goodwin. 2017. Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review* 21, 2 (2017), 99–141.

[39] Sofiane Ramdani, Benoît Seigle, Julien Lagarde, Frédéric Bouchara, and Pierre Louis Bernard. 2009. On the use of sample entropy to analyze human postural sway data. *Medical Engineering and Physics* 8 (2009), 1023–1031.

[40] Tehran J. Davis Michael A. Riley Kevin Shockley Ramenzoni, Veronica C. and Aimee A. Baker. 2011. Joint action in a cooperative precision task: Nested processes of intrapersonal and interpersonal coordination. *Experimental Brain Research* 211 (2011), 447–457.

[41] Ellen Reuzel, Petri JCM Embregts, Anna MT Bosman, Ralf Cox, Maroesjka van Nieuwenhuijzen, and Andrew Jahoda. 2013. Conversational synchronization in naturally occurring settings: A recurrence-based analysis of gaze directions and speech rhythms of staff and clients with intellectual disability. *Journal of nonverbal behavior* 37, 4 (2013), 281–305.

[42] 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.

[43] Daniel C Richardson, Rick Dale, and John M Tomlinson. 2009. Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science* 33, 8 (2009), 1468–1482.

[44] Michael A Riley, Michael Richardson, Kevin Shockley, and Verónica C Ramenzoni. 2011. Interpersonal synergies. *Frontiers in psychology* 2 (2011), 38.

[45] Emanuel A Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in society* 29, 1 (2000), 1–63.

[46] Bertrand Schneider and Roy Pea. 2013. Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-supported collaborative learning* 8, 4 (2013), 375–397.

[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] Julian Schulze and Stefan Krumm. 2017. The "virtual team player" A review and initial model of knowledge, skills, abilities, and other characteristics for virtual collaboration. *Organizational Psychology Review* 7, 1 (2017), 66–95.

[49] Kevin Shockley. 2005. Cross recurrence quantification of interpersonal postural activity. *Tutorials in contemporary nonlinear methods for the behavioral sciences* (2005), 142–177.

[50] Kevin Shockley, Daniel C Richardson, and Rick Dale. 2009. Conversation and coordinative structures. *Topics in Cognitive Science* 1, 2 (2009), 305–319.

[51] Ivan D Steiner. 2007. Group process and productivity (social psychological monograph). (2007).

[52] Angela Stewart and Sidney K D'Mello. 2018. Connecting the Dots Towards Collaborative AIED: Linking Group Makeup to Process to Learning. In *International Conference on Artificial Intelligence in Education*. Springer, 545–556.

[53] Chen Sun. 2019. Toward a Generalized Competency Model of Collaborative Problem Solving. *In review*. (2019).

[54] Michael T Tolston, Kevin Shockley, Michael A Riley, and Michael J Richardson. 2014. Movement constraints on interpersonal coordination and communication. *Journal of Experimental Psychology: Human Perception and Performance* 40 (2014), 1891–1902.

[55] Preethi Vaidyanathan, Jeff Pelz, Cecilia Alm, Pengcheng Shi, and Anne Haake. 2014. Recurrence quantification analysis reveals eye-movement behavior differences between experts and novices. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 303–306.

[56] Sebastian Wallot. 2017. Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes* 54, 5-6 (2017), 382–405.

[57] 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* 7 (2016), 1835.

[58] Jacqueline Kory Westlund, Sidney K D'Mello, and Andrew M Olney. 2015. Motion Tracker: camera-based monitoring of bodily movements using motion silhouettes. *PloS one* 10, 6 (2015), e0130293.

[59] Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://ggplot2.org

[60] Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21, 1 (2017), 173–186.