Author/s:
Pöysä-Tarhonen, J;Li, S;Hautala, J;Awwal, N;Häkkinen, P

Title:
Making the Invisible Visible: Exploring Joint Attention Behaviour in Remote Collaborative Problem-Solving

Date:
2025-08-01

Citation:
Pöysä-Tarhonen, J., Li, S., Hautala, J., Awwal, N. & Häkkinen, P. (2025). Making the Invisible Visible: Exploring Joint Attention Behaviour in Remote Collaborative Problem-Solving. Journal of Computer Assisted Learning (JCAL), 41 (4), https://doi.org/10.1111/jcal.70068.

Persistent Link:
https://hdl.handle.net/11343/362520

Journal of Computer Assisted Learning

WILEY

Journal of Computer Assisted Learning

ORIGINAL ARTICLE OPEN ACCESS

Making the Invisible Visible: Exploring Joint Attention Behaviour in Remote Collaborative Problem-Solving

Johanna Pöysä-Tarhonen[1] | Shupin Li[1] | Jarkko Hautala[2] | Nafisa Awwal[3] | Päivi Häkkinen[1]

[1]Finnish Institute for Educational Research, University of Jyvaskyla, Jyväskylä, Finland | [2]Niilo Mäki Institute, Jyväskylä, Finland | [3]Assessment and Evaluation Research Centre, University of Melbourne, Melbourne, Australia

Correspondence: Johanna Pöysä-Tarhonen (johanna.poysa-tarhonen@jyu.fi)

Received: 23 August 2024 | **Revised:** 16 April 2025 | **Accepted:** 15 May 2025

Funding: This work was supported by Academy of Finland, 273970.

Keywords: collaborative problem solving | cross-recurrence quantification analysis | eye-tracking | joint attention | qualitative analysis

ABSTRACT

Background: Collaborative problem-solving (CPS), encompassing social and cognitive domains, is recognised as a critical competency for current and future learners. While the cognitive domain, rooted in individual problem-solving, is relatively well understood, the social domain, manifested through social interaction, remains a challenge, especially in remote settings.

Objectives: This study explored the complex social domain of CPS through theoretical and empirical understanding of joint attention (JA) behaviour. JA was conceptualised as a continuum of attentional levels, from lower (monitoring and common) to higher levels (mutual and shared). While gaze alignment was considered foundational to JA, capturing higher attentional levels, crucial for productive CPS, required extending the focus to communication over the shared objects of attention—were they external (e.g., events) or internal (e.g., thoughts).

Methods: Dual eye-tracking data from 12 dyads (aged 12–13) engaged in remote CPS were first analysed for the strength of eye-gaze coupling, used as a proxy for lower JA levels and a potential indicator of higher levels. This was followed by qualitative analysis of interaction (logfiles, eye-event videos) in three dyads with strong eye-gaze coupling, to examine how these measures related to actualised JA.

Results and Conclusions: The results highlighted the relevance of JA in understanding the social domain of CPS, even in remote settings. Informed by quantitative measures, the qualitative analysis revealed dyadic behaviour linked to higher JA levels. Overall, the findings provide preliminary insights into how JA relates to productive CPS, offering a groundwork for further research on supporting learners to "improve collectively".

1 | Introduction

In response to the multifaceted social, economic, environmental, and technological challenges of the contemporary world, collaborative problem-solving (CPS) is increasingly recognised as one of the key competencies for both current and future learners. Accumulating evidence suggests that individual and CPS differ significantly in both their processes and outcomes (Sears and Reagin 2013). Whereas individual problem-solving involves a single person independently identifying, analysing, and resolving a problem without the input of others (e.g., Chi et al. 1982), CPS typically involves the coordinated engagement of two or more individuals, working toward a shared goal—one that cannot be achieved by a single participant alone (Graesser et al. 2020). While a division of labour exists, contributions are interdependent and inherently interwoven (Care et al. 2016).

Summary

- What is already known about this topic
  - To collaboratively solve problems on society-wide concerns is recognised as among the key competences required by current and future learners, taking also a notable role in recent educational research.
  - While in collaborative problem-solving (CPS), the problem-solving part, with its origin in individual problem-solving, is easier to understand, the collaborative part might remain more challenging to capture, particularly in the remote interaction context.
  - Joint attention (JA) behaviour, the very basis of social interaction predicting productive collaboration, is considered a promising conceptual tool for uncovering collaborative processes within CPS.
- What this paper adds
  - To study JA, this study extends from earlier work that often focuses on studying joint visual attention (JVA) in CPS situations. This study adopts a more nuanced perspective that accounts for varying levels and a richer understanding of JA.
  - The study utilises a rich set of data and methods, all applied in a challenging remote interaction context. Gaze data enhances understanding of remote CPS by going beyond the interaction sequences and making the invisible social moments more visible in partners' interaction but requires careful consideration in bridging theory and data.
- Implication for practice and/or policy
  - The paper stimulates thinking about how to support learners to improve collectively in terms of what contributes to achieving higher attentional levels and thereby more productive collaborative learning and CPS.

As such, the complex CPS construct encompasses both the task (cognitive domain) and the social infrastructure (social domain) through which participants create and share knowledge, monitor their progress, and identify and resolve communication breakdowns (Alterman and Harsch 2017; Roschelle and Teasley 1995). This paper focuses on the social domain, studied in a remote CPS context involving student dyads. While the cognitive domain, rooted in individual problem-solving approaches, is more easily traceable through pre-defined problem-solving paths, the dynamic nature of the social domain, manifested through social interaction, is considerably more challenging to capture (Funke et al. 2018). This challenge is particularly evident in remote interaction settings—the context of this study—where communication is inherently limited compared to face-to-face situations.

How, then, can we achieve a better understanding of the social domain of remote CPS? Digital environments enable the collection of rich process data (e.g., logfiles), consisting of time-stamped sequences of events (e.g., actions and chat), which can be methodologically treated as time series describing the actions and interactions between collaborators. In CPS, these sequences are considered indicators of underlying cognitive and social processes (Care et al. 2016). However, regarding the social domain of CPS, major challenges lie in determining the actual meaning of these sequences and accurately interpreting the underlying social processes between partners (Vista et al. 2016).

Within this context, a promising approach to unveil the social domain of CPS is through the theoretical and empirical understanding of JA behaviour (e.g., Carpenter and Liebal 2012; Mundy 2018; O'Madagain and Tomasello 2019; Siposova and Carpenter 2019; Tomasello 1995). Focusing on JA behaviour in CPS is highly relevant, as JA is integral to successful CPS (e.g., Schneider and Pea 2013; Schneider et al. 2018). Without achieving JA, partners are less likely to establish common ground (e.g., Baker et al. 1999; Clark and Brennan 1991), take the other's perspectives (e.g., Moll and Meltzoff 2012), and collaboratively build on ideas to solve problems (e.g., Barron 2003).

Commonly, JA is defined as a capacity to focus together with another on an external source or object in the environment (e.g., Mundy 2018). Gaze plays a central role in this process, functioning as a primary observable, non-verbal indicator of JA behaviour. This emphasis on gaze is grounded in a foundational assumption within cognitive science—namely, that eye tracking provides a window into the mind, often referred to as the eye–mind link (Just and Carpenter 1980). In collaborative situations such as CPS, eye-tracking methods have been used to trace JA behaviour by extracting measures of how partners focus on a common reference and monitor one another's attention to an external entity (e.g., object, person, or event) during a shared task; commonly referred to as *joint visual attention* (JVA) (e.g., Olsen et al. 2017; Schneider and Pea 2013, 2014; Schneider et al. 2018; Wisiecka et al. 2023). Prior research has shown that visibility of a partner's gaze increases instances of JVA and correlates with higher quality collaboration (Schneider and Pea 2013, 2014). Similar findings have been found across different eye-tracking settings, including co-located collaborations, where increased JVA measures were associated with improved task performance and learning outcomes (Schneider et al. 2018). While the eye-tracking research on collaboration and CPS has largely focused on JVA, this paper extends from this approach and takes a more nuanced approach by accounting for varying levels of JA behaviour. It is argued that although gaze alignment provides a foundational basis for JA, productive CPS involves more—it also requires intentional communication and coordination over shared objects of attention. In CPS, the objects of attention often go beyond visual or auditory stimuli to encompass temporally extended or mental contents such as present, past, and future events or ideas, plans, and reasons (O'Madagain and Tomasello 2019). Thus, while achieving JA to external contents is considered a perceptual phenomenon (e.g., JVA), JA directed at mental contents, is primarily reflected in the linguistic exchanges (O'Madagain and Tomasello 2019). In contrast to this "richer" account of JA observed in productive CPS, behavioural indicators such as gaze following and referential looking represent more of a "minimal" account (León 2022; Wilby 2023) yet serve as a valuable foundation for further analyses—particularly when combined with qualitative approaches (Schneider and Pea 2014). Accordingly, this paper moves beyond the minimal account of JA as mere gaze following or referential looking to encompass richer accounts during remote CPS, including considering dyads' intentional interaction, integral to productive CPS.

The subsequent section delves into a more detailed description of the concept of JA behaviour, comprising various levels or intensities and their behavioural indicators.

1.1 | Theoretical Construct: 'Typology of Jointness'

To interconnect the various accounts of JA behaviour, Siposova and Carpenter (2019) propose a sliding scale which connects the diverse intensities of JA, all of which are referred to as JA in the related literature. They argue that, instead of a single state or binary event, JA can be considered a process comprising various, hierarchically nested and closely connected but distinct phenomena. (Siposova and Carpenter 2019, 261) propose a 'typology of jointness' that comprises four attentional levels: *monitoring*, *common*, *mutual*, and *shared*, all of which include the notion of a triadic relationship between self, other, and an object of attention (Figure 1). These levels are distinctive in terms of (1) the participant's perspective (second- and third-person perspectives; Moore and Barresi 2017), (2) the type of knowledge related to a particular attentional level and (3) the level of dependency between partners and level of experience (individually or jointly created). An essential precondition for each of the four levels is the individual's ability to engage in individual attention. This refers to a situation in which an individual is attending to something in the environment with a first-person perspective. In contrast to the other levels in the scale of jointness (from monitoring to sharing of attention), this type of interaction is not triadic but dyadic[1] (a relationship between self and an object of attention), and the type of knowledge level is individual.

1.2 | Indicators of JA at Different Levels

The first level, monitoring attention (Siposova and Carpenter 2019), refers to a situation in which an individual adopts an observer's perspective on a second individual, attending to the same object or situation as the partner. At this level, the participants hold individual knowledge of the situation, and their attentional levels are independent, although each is aware that the other is also attending to the same object or situation. While both individuals simultaneously monitor each other's attention to the object or situation, they still assess the attention and knowledge states of the other participant individually. Whereas monitoring behaviour is often observable through gaze or bodily orientation, it can also manifest more subtly. The knowledge involved at this level is individual.

At the second level, common attention, two individuals adopt an observer's perspective and attend—nearly simultaneously—to what the other is focused on (Siposova and Carpenter 2019). They not only attend to the same object or situation but also to each other's attention toward it. Engaging in common attention requires the object to be clearly pronounced and marked, allowing participants to assume that they are attending to the same target. Additionally, they must have a reason to monitor the other's' attention, such as a predefined shared goal that makes it relevant. While both are aware they are engaged in the same situation, the evaluation of whether they are in common attention remains individual—and may be inaccurate. The knowledge level is common.

According to Siposova and Carpenter (2019), at the third and fourth levels—*mutual* and *shared*—participants no longer adopt an observer's (third-person) perspective. Instead, the experience is grounded in direct commitment to one other, with both acting as senders and receivers of information. Through this reciprocal interaction, each becomes a constituent part of the other's attentional experience (Zahavi 2015), and attention to the object or situation is shaped by mutual awareness of each other's focus (Siposova and Carpenter 2019). This bidirectional nature distinguishes these levels from the more individualistic monitoring and common attention (Siposova and Carpenter 2019). In *mutual attention*, participants are aware that their partner is attending both (a) to the same object or situation and (b) to their own attention toward it, even without intentional communication (Siposova and Carpenter 2019). Compared to common attention, the experience is co-created, and the knowledge type is *mutual*.

The fourth level, *shared attention*, builds on mutual attention but also requires participants to deliberately communicate about the object or situation and/or the fact that they are jointly attending to it (Siposova and Carpenter 2019). What distinguishes shared attention is its intentional nature. The experience is co-created and marked by behaviours in which individuals intentionally direct and follow each other's attention to confirm shared focus (Gilbert 2007; Siposova and Carpenter 2019). At this level, the knowledge type is *shared*.

1.3 | Aims

This exploratory study aims to deepen understanding of the social domain of remote CPS by focusing on JA behaviour in student dyads. More broadly, the study seeks to stimulate thinking

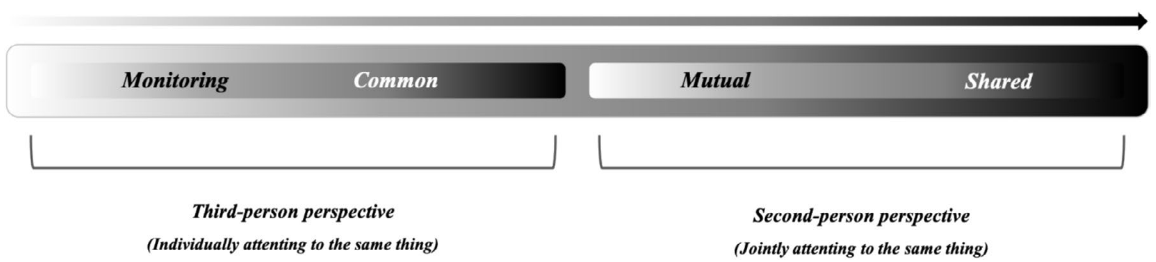


FIGURE 1 | Scale of jointness in joint attention behaviour. Modified from Siposova and Carpenter (2019).

on how to support students achieving higher attentional levels, and in turn, more productive CPS.

The study relies on the theoretical understanding of JA as a sliding scale of jointness (Siposova and Carpenter 2019) and positions JA as integral to productive CPS. The study adopts a multi-method approach, combining successive quantitative and qualitative phases, and draws on compound evidence from multimodal data. First, quantitative indicators of JA are identified in students' gaze behaviour during remote CPS, with eye-gaze coupling serving as a proxy for lower JA levels and a potential indicator of higher levels. Second, based on these measures, three dyads—particularly those with strong eye-gaze coupling—are selected for qualitative analysis of interaction to interpret the quantitative measures in relation to the actualised JA behaviour. Additionally, symmetry of partners' eye-gazing, indicating possible leader-follower roles in the dyads, is examined in the selected cases.

The following research questions are posed:

RQ1. *What is the strength of dyads' eye-gaze coupling during remote CPS processes? Is the coupling symmetrical or do the dyads spontaneously form leader–follower roles?*

RQ2. *What do these eye-gaze measures indicate about actualised levels of JA behaviour in selected dyads, and do higher values correspond to higher levels of JA attention at the interaction level?*

2 | Materials and Methods

2.1 | Participants and Procedure

The research participants ($n = 24$) were 12–13-year-old students (*M*age 12.53) from schools in a Finnish urban municipality. Participation was voluntary, with informed consent obtained from both students and their guardians. Ethical approval was granted by the university's ethics committee.

The data were collected in a dual eye-tracking setting (Olsen et al. 2017) as participants completed a bundle of CPS tasks within a digital CPS environment. The students were randomly assigned into 12 dyads and completed the tasks remotely in separate cognitive labs.

2.2 | CPS Environment and Tasks

The digital CPS environment[2] features a graphical interface presenting a predefined problem space, in which dyads interact with both the task and each other (Awwal et al. 2015). It includes a free-form chat for synchronous communication, enabling dyads to develop their own communication strategies, and a space with actionable artefacts, which may appear symmetrically or asymmetrically. In symmetrical tasks, both partners receive identical stimuli content and artefacts; in asymmetrical ones, each receives a unique subset of resources or interface controls. In some cases, the screen view remains identical but actions—such as object manipulation or scrolling—are restricted to one partner. The majority of tasks are deliberately asymmetrical by design to promote collaboration through interdependence and information sharing. Moreover, dyads could progress through the environment without necessarily reaching a correct solution. In the tasks, one learner's success depends on their partner's behaviour and responses (Griffin and Awwal 2018).

This paper focuses on the Plant Growth task (Figure 2, Awwal et al. 2015), one of three pre-bundled, asymmetric CPS tasks (*Plant Growth, Olive Oil and Game of Twenty*). While the students were familiar with the environment from a prior data collection session held a year earlier, the Plant Growth task was new to them. The task unfolds over two levels (Pages 1 and 2) and provides each participant with asymmetric resources. On Page 1, Student A controls light density and Student B controls temperature, with each only able to observe, not manipulate, their partner's controls. Together, they must identify consistent patterns in plant growth across combinations of light (from very dark to very bright) and temperature (10°C C–40°C). This process requires active information exchange to infer each partner's control mechanism. Participants are prompted to observe, collect data, identify patterns, form and test rules, generalise findings, and make hypotheses. Page 2 builds on this by asking students to apply their understanding to determine optimal or minimal growth conditions and place the plant accordingly on a grid. The first level functions as an exploration phase, while the second assesses rule transfer and application. Task-specific CPS elements are summarised in Table 1.

2.3 | Data

2.3.1 | Eye-Tracking Data and Pre-Processing

During CPS task completion, participants' eye movements were recorded with desktop eye trackers (screen based; SensoMotoric Instruments [SMI] RED 250 hz Mobile). The stimuli were presented on an HP Zbook 15 G2 laptop (15.6-in. display) with a 1920 × 1080 resolution, and a chin rest was installed at a 60-cm viewing distance. A (13-point) calibration was conducted prior to the experiment and before each task.

Eye-tracking data pre-processing was conducted with Behavioural and Gaze Analysis (SMI BeGaze). Saccades were detected with a minimum eye angular velocity parameter value of 40 deg/s. The eye tracker also recorded screen capture videos while the participant worked in the CPS. This video data enabled segmentation of the recordings into CPS subtasks by setting custom trial time periods on a tens of millisecond scale. Then, the corresponding area-of-interest (AOI) definition file was uploaded on each segment, allowing each fixation to be assigned to task-relevant screen objects (for the AOIs on Pages 1 and 2, see Figure 3). In addition, scan path videos for each participant were exported to enable qualitative analysis. In scan paths, the gaze positions and eye events were plotted on a stimulus video.

Next, fixation data was exported to be further processed in the SPSS 27 (IBM) program. Fixation data for each segment was transformed into time series data that assigned an AOI value for each second starting from the beginning of the segment. The

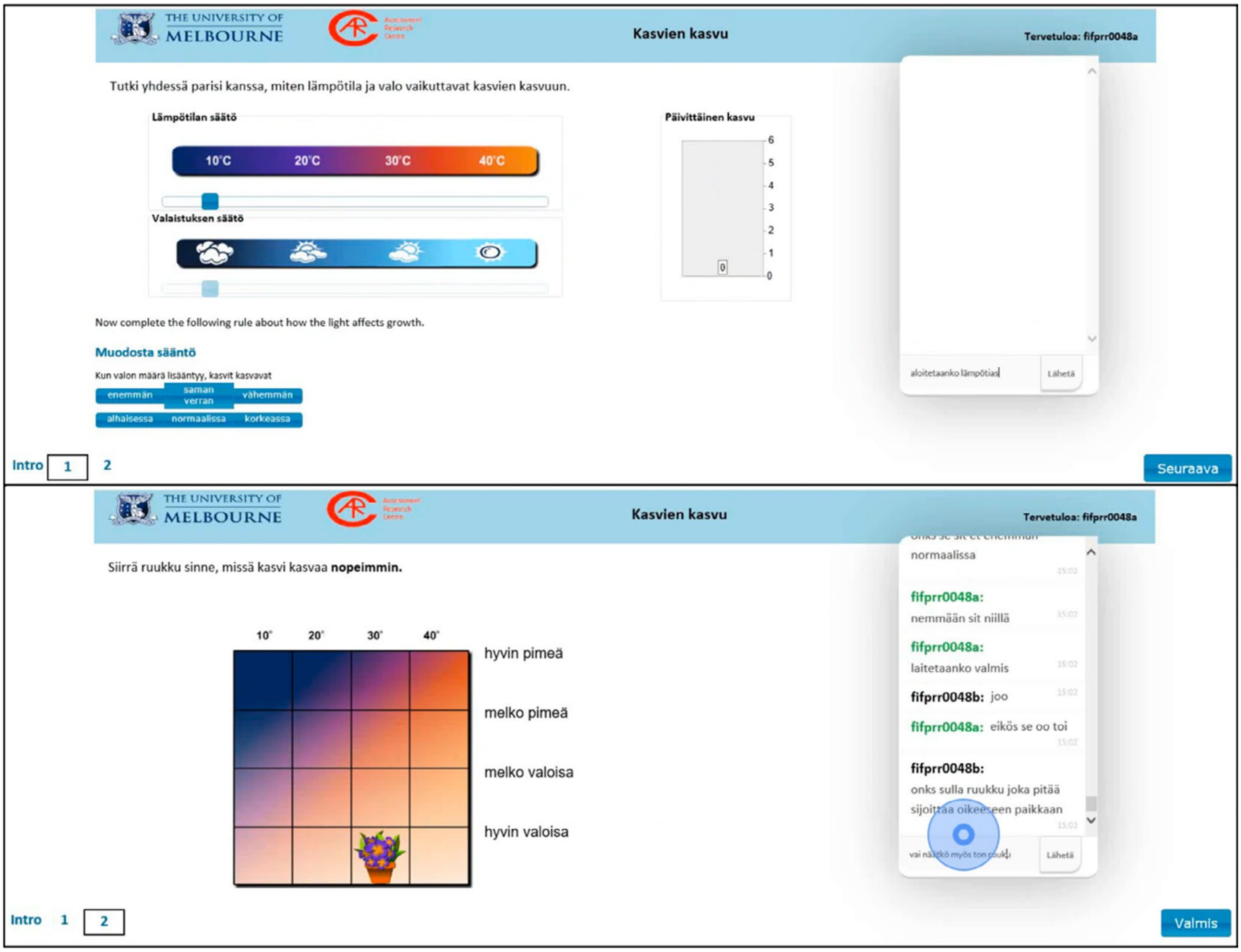


FIGURE 2 | Screen captures from the Plant Growth task (Pages 1 and 2). Student A's display screen (in Finnish).

TABLE 1 | Task-specific elements regarding the CPS construct (Plant Growth task).

Skills	Behaviour	An example of data captured for assessment
Interaction	Interacting with partner.	Presence of chat during actions and processes.
Responsibility Initiative	Taking responsibility for progress for parts of the group task.	Realising the need to move to subsequent pages of the task and communicating with partner.
Responsiveness	Responding to the contributions of others.	Responding to partner's specific queries before proceeding with other activities.
Collecting Information	Recognising the need for more information.	Undertaking activities with relevant and available resources.
Systematicity	Implementing possible solutions to a problem.	Trial of different combinations of light density and temperature options in search of plausible conditions.
Solution	Answering correctly.	Placement of the plant in the correct position on the grid by Students A and B, respectively, as per the conditions given.

Note: Adapted from (Awwal et al. 2015).

AOI area receiving the most fixation time during that second was assigned. Missing values resulting from eyelid closures or looking away from the screen were kept in the data, as were fixations on non-AOI areas.

2.3.2 | Logfiles

The logfiles captured chat interactions and individual or joint on-screen activity, including some non-activities (e.g., mouse

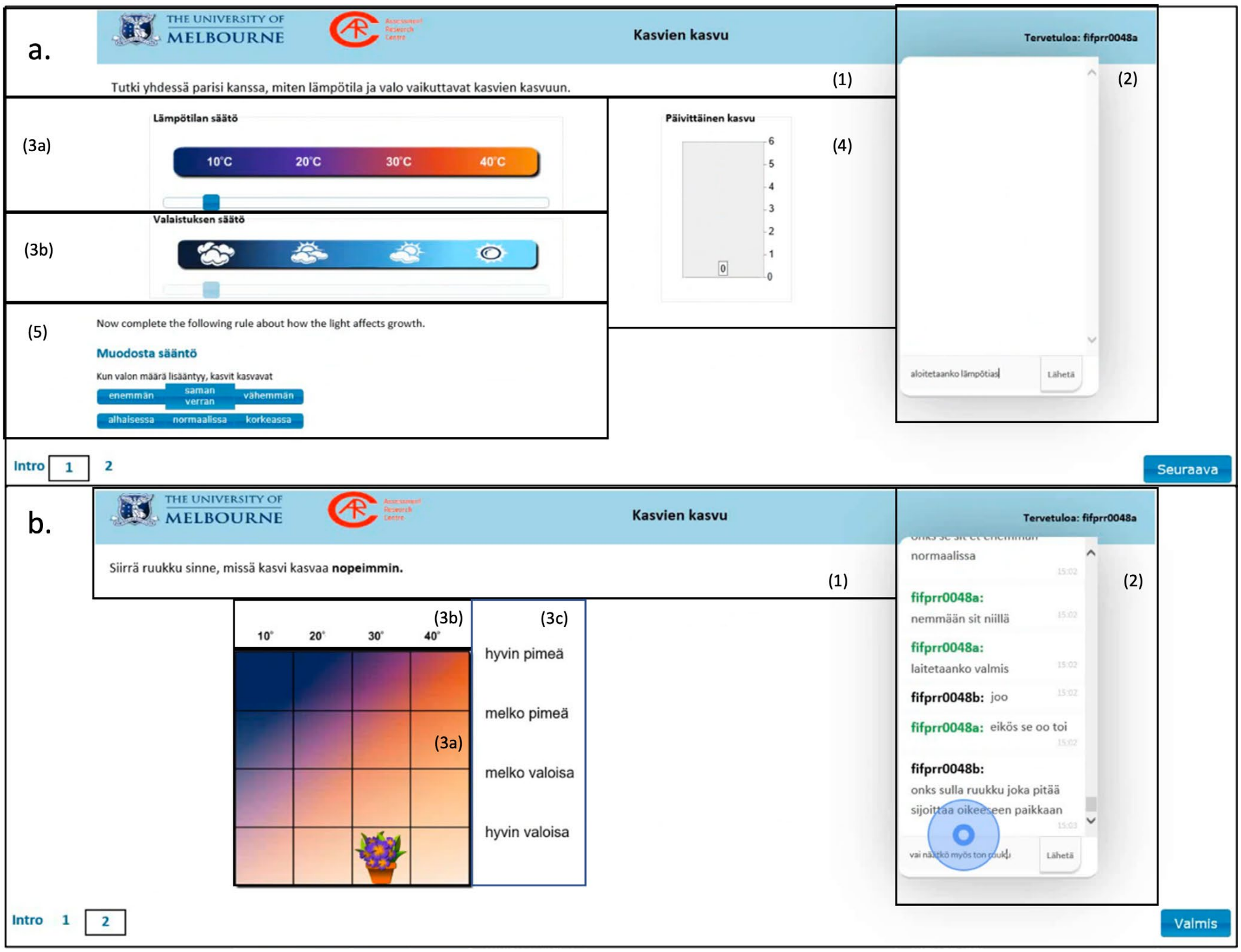


FIGURE 3 | Areas of interest for the plant growth task. The defined AOIs are numbered as 1–5 (Page 1) and as 1–3c (Page 2). Student A's view (in Finnish).

movements, hovering over a button without clicking). All captured information was recorded as a sequence of activities in the order in which they occurred, including the time of the occurrence, the details of the involved participants, and the task they were undertaking (including task page numbers).

2.4 | Data Analysis

Data analysis included two successive phases as quantitative and qualitative (Table 2). First, the strength of dyads' eye-gaze coupling during CPS was measured using Cross-Recurrence Quantification Analysis (CRQA), with the measures serving as a proxy of lower JA levels and a potential indicator of higher levels. Second, these measures guided the selection of dyads for further analysis of DCRP, followed by qualitative analysis of interaction through logfiles and eye-event videos to elucidate the quantitative measures against the actualised JA levels.

2.4.1 | Quantitative Analysis of Eye-Gazing

The strength of eye gaze coupling within each dyad was examined using CRQA (e.g., Coco and Dale 2014; Coco et al. 2020; Marwan et al. 2007; Wallot and Leonardi 2018). CRQA, which builds on the cross-recurrence plots (Wallot and Leonardi 2018; see Appendix 1 for an example), quantifies the temporal similarity between two time series (Coco et al. 2020) and enables visualisation of their temporal evolution through CRPs (Marwan et al. 2007). In dual eye-tracking data, the time series represent the fixation sequences of two collaborators, including the x- and y-coordinates of fixations and their respective timestamps (Marwan et al. 2007). To move beyond the visual impression of the sequential similarity (Wallot and Leonardi 2018), this study focused on CRQA metrics that quantify the small-scale structures within the recurrence plots (Marwan et al. 2007), with particular *emphasis on recurrence rate* (*RR*) *as the central measure of CRQA*.

CRQA was conducted using the CRQA package (v 2.0.2; Coco and Dale 2014) in R 4.1.0 (R Core Team 2021). Categorical time series data were used to represent each participant's gaze sequences across specific AOIs on the two pages of the Plant Growth task (Page 1 included six AOIs, Page 2 five; Figure 3). In total, 48 separate time series were generated for 24 participants across the two pages. Prior to applying CRQA, missing data—that is, instances where gaze was directed outside the screen—were removed, along with the corresponding data points of their partners. After

TABLE 2 | The different phases of the multi-method data analysis on eye-tracking and logfile data.

Phase of the analysis	Data source/level	Analysis	Target, expected output
Phase 1 (Quantitative)	ET data (categorial time series data)	CRQA	Strength of coupling of students' eye-gazing.
	ET data of selected dyads	DCRP	Symmetry of coupling of students' eye-gazing.
Phase 2 (Qualitative)	Log data of selected dyads	Coding the initiating and responsive actions on the logfiles.	Sequential characteristics of interaction as meaningful events related to attentional levels of JA and quality of CPS process.
		Identifying CPS elements from the coded data.	Contextualisation of the coded events.
	Eye-event videos of selected dyads	Coding the location and order of the eye-events at specific AOIs during the selected events (to analyse for congruence; frame-by-frame analysis).	Identification whether and how gaze data can reinforce or modify the initial interpretation of interaction, based on logfiles.

preprocessing, the number of valid data points ranged from 103 to 1180 on Page 1 and from 52 to 454 on Page 2, which was sufficient for performing CRQA (Wallot and Leonardi 2018). For the analysis, the following parameters were used for the categorical data: *delay* = 1, *embed* = 1, *radius* = 0.0001, and *rescale* = 0 (Wallot and Leonardi 2018).

2.4.1.1 | CRQA Measures. CRQA measures are derived from the *recurrence point density* and the *diagonal* and *vertical line structures of CRP* (e.g., Coco and Dale 2014; Coco et al. 2020; Marwan et al. 2007) (Table 3). For the diagonal dimensions, the following measures were calculated: *RR*, *determinism* (*DET*), the *average diagonal length* (*L*), the *length of the longest diagonal* (*Lmax*), and the *entropy* of the diagonal line length distribution (*ENTR*, Shannon entropy). For the vertical dimensions, *laminarity rate* (*LAM*) and *trapping time* (*TT*) were calculated. RR values served as the basis for selecting three dyads for further analysis. The selection was based on the highest means of RRs, calculated by combining the values from both pages of the task.

2.4.1.2 | Diagonal Cross-Recurrence Profiles. As a supplementary analysis, DCRPs were calculated for the selected dyads. The DCRP is a follow-up analysis to CRP (Wallot and Leonardi 2018), computing RRs across a limited set of diagonals (a limited range of lags) within a window centred on lag 0, corresponding to the Line of Synchrony (LoS). This approach accounts for the possibility that cross-recurrences in the time series may not align precisely with the central diagonal of the CRP, but may instead be asymmetrically distributed around it (Wallot and Leonardi 2018). The asymmetry can be used to estimate *leader-follower dynamics*, as visualised through *DCRPs* (see Appendix 2). When applying DCRP to categorical time series, the use of default cross-recurrence parameters is recommended (i.e., *delay* = 1, *embedding* = 1 and *radius* = 0; Wallot and Leonardi 2018). In this study, 41 lags were examined (20 on each side of the 33 LoS plus the LoS itself as lag 0).

2.4.2 | Qualitative Analysis of Interactional Episodes

The qualitative analysis of interaction in the selected dyads drew on logfiles at the dyad level and eye-event videos at the individual level (see Table 2). To identify meaningful episodes related to JA and the quality of CPS, the analysis focused on the structure of the interaction and the relationship between interactional moves. The data for each dyad were analysed—including identification, coding, and interpretation—by two researchers. Observations were compared to minimise subjectivity and to reach a shared understanding of their meaning. At this stage, DCRPs were also examined in relation to interaction quality, with particular attention to potential leader–follower dynamics within the selected dyads.

2.4.2.1 | Analysis on Logfiles. The logfile analysis was grounded on the fundamental principles of interaction structure and organisation, which were further refined to examine coordination and mutual understanding in CPS within a dual eye-tracking setting (Andrist et al. 2015, 2018). Interactional actions are typically organised as sequences, often taking the form of adjacency pairs. For instance, an *initiating action* (e.g., a question or proposal) creates an expectation for a *responsive action* (e.g., an answer or uptake) in the sequential position. In triadic interactions involving JA, one participant may initiate interaction by directing the other's attention to a particular object; in this context, the referent is often an artefact within the collaborative workspace. Within the CPS environment used in this study, the objects of attention may be abstract (e.g., numerical problems) or concrete (e.g., on-screen artefacts) that are explicitly present and manipulable within the task interface.

In the line-by-line manual coding, initiating and responsive actions included verbal exchanges (chat), as well as combinations of chat and action (e.g., when one student prompts their partner to take some action and subsequently provide feedback on the outcomes). To enable more detailed categorisation of interactional events, a *reference-action sequence* was applied (Andrist

TABLE 3 | CRQA measures.

CRQA measures	Definition	Definition in eye-tracking data
Diagonal		
Recurrence rate (RR)	Density of recurrence points in CRP. Higher density of RRs in diagonal result in higher RR value.	Percentage of cross-recurrent fixations.
Determinism (DET)	Indicates the proportion of recurrent points forming diagonal lines in CRP. Stochastic, uncorrelated behaviour and highly fluctuating time series cause no or short diagonals; deterministic processes lead to longer diagonals and less isolated recurrence points.	Percentage of identical scan path segments.
Average diagonal length (L)	Average time both trajectories are close to each other. High DET and L values suggest consistent similar dynamics in both trajectories.	Average time partners are exhibiting similar gaze patterns.
Length of the longest diagonal (Lmax)	Indicates the longest period the time series stay attuned.	Longest period in which the scan paths are synchronised
Entropy of the diagonal line length distribution (ENTR, Shannon entropy)	Complexity of diagonal lines (distribution of recurrence points in CRP). Higher values indicate more scattered time series; lower values more synchronised time series.	Complexity of relation between scan paths.
Vertical		
Laminarity rate (LAM)	Vertical lines in CRP. Higher rate suggests greater stability in gaze behaviour, lower rates indicate more variability.	The extent to which partners maintain consistent patterns of gaze alignment over time.
Trapping-time (TT)	Average length of vertical structures. Indicates the average time two trajectories stay in the same region.	Prolonged time during which partners focus on certain AOI.

Note: Based on (Coco and Dale 2014; Coco et al. 2020; Marwan et al. 2007; Villamor and Rodrigo 2019).

et al. 2018). These sequences represent "short interactions between collaborators in which one person indicates an object in the collaborative workspace that another person is supposed to manipulate in some way" (Andrist et al. 2018, p. 339). In the current study, such sequences were initially coded as *initiating-responding* but were later redefined as *reference-action* sequences for greater clarity (Table 4).

The resulting interactional structures were then examined in relation to theorised attentional levels (Siposova and Carpenter 2019). In addition, to contextualise the coded events, the process qualities of CPS interaction were identified. Specifically, behavioural indicators of task-specific CPS elements—such as responsibility, initiative, responsiveness, collecting information (Awwal et al. 2015)—were identified from the logfiles (see Table 1).

2.4.2.2 | Qualitative Analysis on Scan Path Videos. Next, the interactional events in the logfiles that were related to higher JA levels were further analysed in relation to the eye-events in the scan path videos (individual level; Table 4). In the frame-by-frame analysis, the location and order of the eye-events regarding the specific AOIs were coded for each selected event. The aim was to determine whether and how gaze data supported or altered the initial interpretation of the interaction observed in the logfiles. Table 4 demonstrates how the analyses of the logfile and scan path video were integrated and mutually enriched. It presents a concrete example of how an episode was analysed by combining interactional data from the logfiles with gaze behaviour from the scan path videos, further contextualised by the CPS elements identified.

3 | Results

3.1 | What Is the Strength of Dyads' Eye-Gaze Coupling During Remote CPS Processes; Is the Coupling Symmetrical? (RQ1)

3.1.1 | Outcomes of the CRQA

Next, the results of the descriptive statistics (means, SDs) of the eye-tracking data and CRQA measures across two pages of the Plant Growth task are presented in Tables 5 and 6. The interpretation of CRQA measures is context dependent

TABLE 4 | Example of a coded log data episode, with corresponding gaze behaviour on the areas of interest as well as observable CPS elements.

Time	Role	Event	Logfile	Code	Sub-code	Scan path AOIs	CPS elements
18.36	B	Chat	Hi	Initiating		2	Interaction
18.36	A	Chat	Can you see temperature and cloudiness rate on the left and at the bottom selection buttons?	Initiating		2, 3a, 3b, 5 2, 3a, 3b, 5	Collecting information
18.36	B	Move resource	Brightness	Responding	Reference action	2, 3b, 5	Action
18.36	B	Chat	Yep			2	Interaction
18.36	B	Chat	Can you adjust lighting?	Initiating		2, 3b, 3a	Collecting information
18.37	A	Chat	And next to the chat box the "daily growth"–box?	Initiating		2, 4	Collecting information
18.37	A	Chat	I'll try	Responding		2, 3a	Interaction
18.37	B	Chat	Yep	Responding		2	Interaction

TABLE 5 | CRQA results and time on task (Page 1 of the Plant Growth task).

Dyad	RR (%)	DET (%)	Lmax	L	ENTR	LAM (%)	TT	Time (s)
2	33.93	79.80	27	4.07	1.91	86.73	5.22	569
4	20.91	56.02	13	3.03	1.37	67.31	3.78	160
5	29.78	78.34	35	4.41	2.02	89.18	6.91	987
6	25.24	70.81	23	3.45	1.64	82.98	5.09	408
7	28.27	71.43	18	3.38	1.60	86.02	5.19	614
8	34.34	84.91	27	4.39	2.03	92.25	6.71	279
9	22.52	76.20	32	4.09	1.89	88.11	6.14	1180
10	23.21	62.27	13	3.60	1.56	83.51	4.45	103
11	29.02	76.14	23	4.19	1.85	82.78	5.19	253
12	31.39	76.39	20	4.72	2.06	80.66	6.57	225
13	32.10	75.87	20	3.89	1.82	90.73	6.00	348
14	19.02	58.96	12	3.16	1.46	73.19	3.96	262
Mean	27.48	72.26	21.92	3.87	1.77	83.62	5.43	449.00
SD	5.19	8.83	7.42	0.54	0.23	7.24	1.05	334.80
High ≥	32.67	81.09	29.34	4.41	2	90.86	6.48	
Low ≤	22.29	63.43	14.5	3.33	1.54	76.38	4.38	

Note: "Time(s)" indicates time spent on Page 1. "High ≥" and "Low ≤" indicate the threshold values for high and low CRQA values.

and might vary across disciplines. According to Villamor and Rodrigo (2019), in the learning sciences, a CRQA value is considered high if it is equal to or exceeds the mean plus one standard deviation, and low if it is equal to or exceeds the mean minus one standard deviation. While the variables of CRQA are distinct, the measures are largely interrelated and depend on the RR. In terms of interpretation, a relatively high RR indicates that many individual instances are recurrent (Wallot and Leonardi 2018).

Based on the highest RR values combined across the two pages of the CPS task (see Table 7), three dyads (Dyads 2, 5, and 8) were

TABLE 6 | CRQA results and time on task (Page 2 of the Plant Growth task).

Dyad	RR (%)	DET (%)	Lmax	L	ENTR	LAM (%)	TT	Time (s)
2	44.96	85.37	24	4.56	2.07	89.18	7.46	215
4	28.07	58.06	9	2.69	1.08	65.93	2.97	114
5	45.40	90.99	31	5.26	2.27	95.63	9.46	454
6	31.65	80.72	10	3.40	1.54	90.03	5.73	54
7	35.05	72.53	11	3.04	1.40	85.64	4.04	118
8	43.33	89.14	20	5.00	2.16	93.52	8.79	181
9	16.04	80.51	36	6.05	2.10	95.22	16.04	345
10	32.62	70.08	15	3.13	1.46	87.41	5.25	172
11	41.12	78.63	13	3.63	1.65	81.38	3.90	155
12	30.58	53.81	8	2.85	1.24	61.19	3.69	52
13	38.33	79.32	23	4.79	2.10	92.24	9.01	214
14	43.28	84.21	27	4.84	2.06	89.52	6.38	158
Mean	35.87	76.95	18.92	4.10	1.76	85.57	6.89	186.00
SD	8.68	11.54	9.32	1.11	0.41	11.08	3.64	115.16
High ≥	44.55	88.49	28.24	5.21	2.17	96.65	10.53	
Low ≤	27.19	65.41	9.6	2.99	1.35	74.49	3.25	

Note: "Time(s)" indicates time spent on Page 1. "High ≥" and "Low ≤" indicate the threshold values for high and low CRQA values.

TABLE 7 | Highest means of the recurrence rates in Pages 1 and 2 (Plant Growth).

Dyad	RR % Mean (Page 1 & 2)
2	39.445
5	37.590
8	38.835

selected for further analysis. This involved estimating DCRPs, followed by qualitative analysis of their interactions.

Next, the CRQA measures obtained from Pages 1 and 2 are briefly described for the selected Dyads 2, 5, and 8. On Page 1 (Table 5), focusing on the most interconnected measures of CRQA (RR, DET, *L*, Lmax; Wallot and Leonardi 2018), all three dyads demonstrated relatively high RR values, with Dyad 8 showing the highest RR among them.

Additionally, DET (the percentage of identical scan paths) was high across all dyads, with Dyad 8 recording the highest value. The (L) ranged from average to high, with Dyad 5 slightly exceeding the others. Lmax (the length of the longest diagonal) was also high for all dyads, peaking for Dyad 5. This indicates that Dyad 8 exhibited high and more homogeneous cross-correlation in AOI fixations, while the gaze coupling of Dyad 5 reflected changing trajectories of low and high correlation.

Regarding ENTR (the probability distribution of diagonal line lengths), values ranged from average to high, with Dyad 2 scoring slightly lower than Dyads 5 and 8. Laminarity rate (LAM), representing the percentage of recurrent points forming vertical lines, also ranged from average to high and was highest for Dyad 8. Trapping time (TT), the average length of the vertical lines, was similarly average or high across all three dyads, with the highest value recorded for Dyad 5.

On Page 2 (Table 6), Dyad 5 exhibited the highest values for RR, DET, and L. Dyad 5 also had the highest Lmax value and showed slightly higher entropy (ENTR) of diagonal line length distribution compared to Dyads 2 or 8, suggesting that fixations of Dyad 5 were more distracted. Furthermore, Dyad 5 displayed slightly higher values for both laminarity rate (LAM) and trapping time (TT), indicating that one partner in Dyad 5 consistently focused on AOIs previously gazed at by the other, and that the dyad spent longer periods focusing on specific AOIs compared to Dyads 2 and 8.

3.1.2 | Diagonal Cross-Recurrence Profiles (DCRP) of Dyads 2, 5, and 8

The DCRP analysis revealed potential leader-follower behaviour within Dyads 2, 5, and 8, suggesting asymmetrical patterns in eye-gaze coupling. These findings indicate variation in how individuals led and followed one another's gaze during the collaborative task. On Page 1, Dyad 2 reached its maximum RR (45.76%) when Student A was 3 s ahead. In Dyad 5, the maximum RR (34.39%) occurred when Student B was 10 s ahead, while in Dyad 8, the maximum RR (52.27%) occurred when Student B was 15 s ahead. On Page 2, Dyad 2 showed the maximum RR (51.76%) when Student B was 16 s ahead of Student A. In Dyad 5, the maximum RR (53.76%) was reached when Student B was 2 s ahead.

In Dyad 8, when Student B was 17 s ahead, they achieved the maximum RR (50.61%).

3.2 | What Do the Eye-Gaze Measures Indicate of the Actualised Levels of JA Behaviour? (RQ2)

Next, the results of the qualitative analysis of interaction (Andrist et al. 2015, 2018) for the selected Dyads (2, 5, and 8) are presented through representative episodes, illustrated with data excerpts from Page 1 of the Plant Growth task. The illustration includes simultaneous moments from the logfile, combined with individual-level screen captures from the scan path videos and the coded location and order of the eye-events at specific AOIs (see Figures 4–7).

3.2.1 | The Manifestation of Shared Attention: Dyad 2

Figure 4, an excerpt from the beginning of Dyad 2's Page 1, illustrates the highest JA level, shared attention (Siposova and Carpenter 2019). To attend *with* each other, the partners proceed differently: whereas Student B commences by first greeting Student A to make an attention contact (Gómez 1996), Student A directly poses a question about the partner's view on the task, thereby creating something to focus on together—something "between us"—that forms the basis for second-person relation between the partners (Siposova and Carpenter 2019). In the excerpt, the dialogue continues bi-directionally as the participants explore one other's views and access to the manipulatives of the task.

Over the course of CPS, if focusing on the logfiles only, Student A, who predominantly leads the discussion (see DCRP results), appears to arrive at the solution independently by testing the temperatures and explaining the outcome to Student B. Their CPS activities indicate that Student B primarily engages in acting out rules (i.e., action) with little communication (i.e., interaction) with the partner. Student A's behaviour in the logfiles reflects several key CPS elements, including audience awareness, responsibility initiative, and adaptive responsiveness. These are demonstrated through continuous self-evaluation, aligning understanding with that of the partner, and maintaining focus on the shared goal (Hesse et al. 2015). However, gaze data reveal further insights into Student B's role. Despite limited verbal contributions, Student B is also actively testing temperatures and independently identifying the rule—though without communicating this to the partner. At one point, Student B appears to begin articulating the discovery in the chat but deletes the message and instead replies simply "Yep" (see Figure 5).

3.2.2 | The Manifestation of Shared Attention: Dyad 5

Figure 6 presents an excerpt from Dyad 5, illustrating a shared attention level between the partners (Siposova and Carpenter 2019). Based on the gaze data, at the beginning of the Page 1, both students carefully navigate through the screen view and manipulatives, reading the instructions multiple times. Each applies diverse methods to attract their partner's attention, establishing the foundation for shared attention experience to arise. Whereas Student B first acknowledges the partner's presence with a salutation, Student A commences by directing attention to

Time	Role	Event	Verbal	Eye gazing (AOI)
18.36	B	Chat	*Hi*	2
18.36	A	Chat	*Can you see temperature–and cloudiness rate on the left and at the bottom selection buttons?*	2, 3a, 3b, 5
18.36	B	Move-resource	*Brightness [as a response discovers the lightning control]*	2, 3b, 5, 2
18.36	B	Chat	*Yep*	2
18.36	B	Chat	*Can you adjust lightning?*	2, 3b, 3a
18.37	A	Chat	*and next to the chat box the "daily growth"-box*	2, 4
18.37	A	Chat	*I'll try*	2, 3a
18.37	B	Chat	*Yep*	2
18.37	A	Move-resource	*Temperature*	3a
18.37	A	Chat	*I can adjust light*	2, 3b
18.37	A	Chat	*No I can't*	2, 3b, 3a
18.37	A	Chat	*Just temperature*	2, 3a, 3b, 2
18.38	B	Chat	*Yes, I have it just the other way around*	3a, 3b, 2

Eye gazing (scan path view)

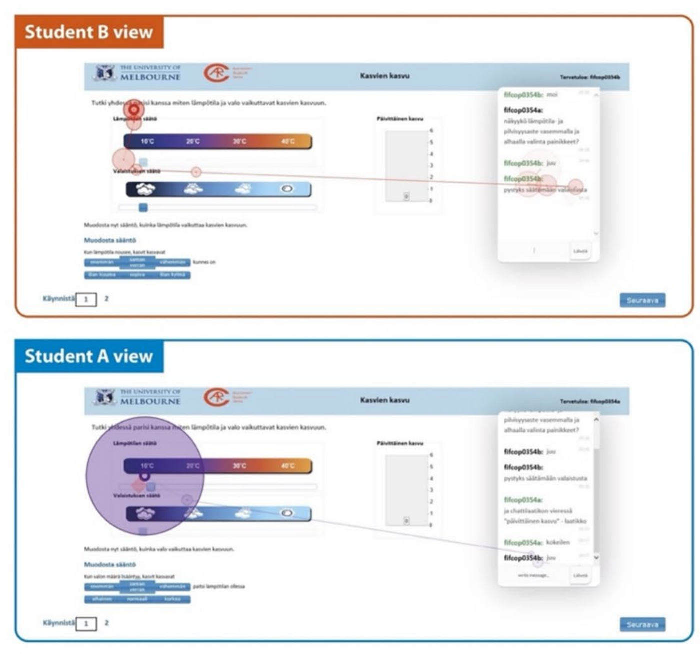


FIGURE 4 | Example from a shared attention-level experience (Dyad 2, Plant Growth). The episodes include logfile data, scan paths and coded eye-event locations at specific AOIs (in Finnish).

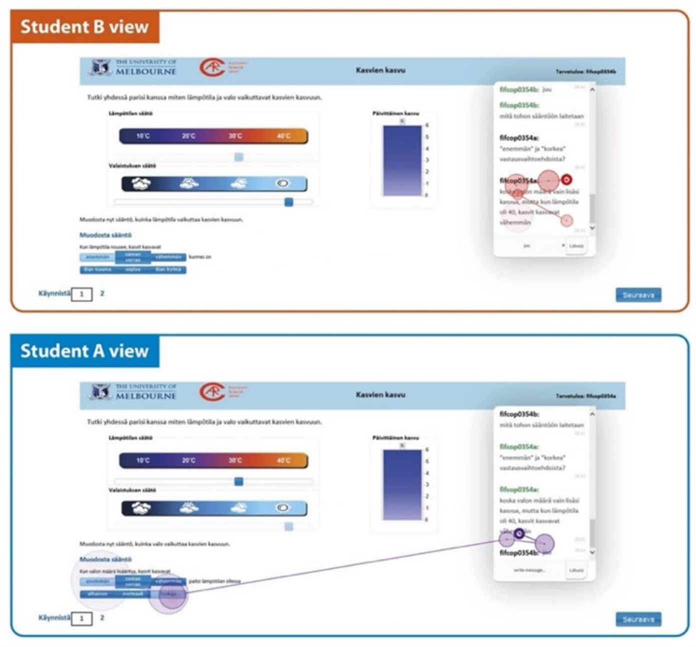
Time	Role	Event	Verbal	Eye gazing (AOI)	Eye gazing (scan path view)
18.41	B	Chat	*What shall we put for the rule there*	2, 5	
18.42	A	Chat	*"More" and "high" of the response alternatives?*	[5, 2, 5] 2	
18.42	B	Rule-confirm	*More,*		
18.42	A	Chat	*Because the amount of light only increased growth, but when the temperature was 40, the plants grow less*	[5, 3b], 2	
18.43	B	Chat	*Yep [starts to write that she/he (B) has different rule, but deletes the text]*	2, 5	

FIGURE 5 | Example from a shared attention-level experience (Dyad 2, Plant Growth). The episodes include logfile data, scan paths, and coded eye-event locations at specific AOIs (in Finnish).

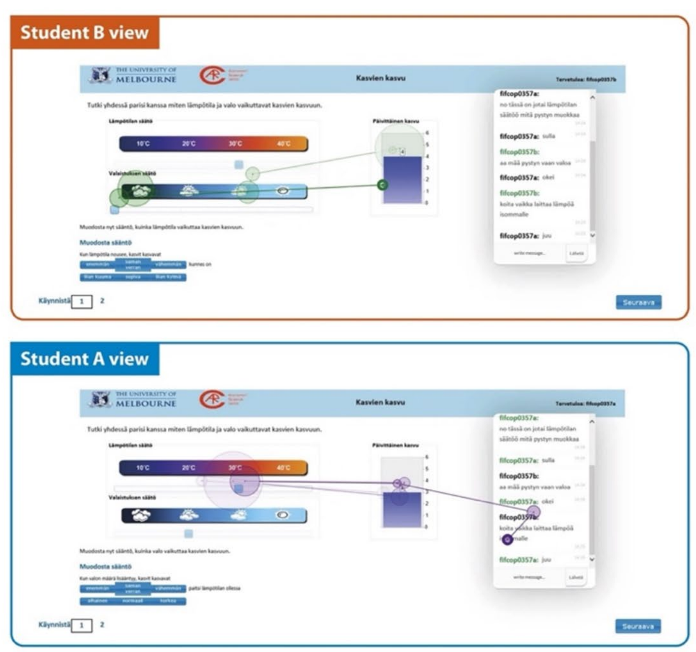
Time	Role	Event	Verbal	Eye gazing (AOI)	Eye gazing (scan path view)
23.24	B	Chat	*Try to set temperature higher for instance*	2	
23.24	A	Chat	*Yep [Starts typing 'If you can adjust lightning, I can...' but deletes the message]*	2	
23.25	A	Move-resource	*Temperature*	3a, 4, 2	
23.25	A	Move-resource	*Temperature*	3a, 4	
23.25	A	Move-resource	*Temperature*	3a, 4	
23.25	A	Move-resource	*Temperature*	3a, 4, 2	
23.25	A	Move-resource	*Temperature*	3a, 4, 2	
23.25	B	Move-resource	*Brightness*	3b, 4	
23.25	B	Move-resource	*Brightness*	3b, 4	
23.25	A	Move-resource	*Temperature*	3a, 4, 5	
23.25	B	Move-resource	*Brightness*	3b, 4	
23.25	B	Move-resource	*Brightness*	3b, 4	
23.25	A	Move-resource	*Temperature*	3a, 4	
23.25	B	Move-resource	*Brightness*	3b, 4	
23.25	A	Move-resource	*Temperature*	3b, 4	

FIGURE 6 | Example from a shared attention-level experience (Dyad 5, Plant Growth). The episodes include logfile data, scan paths, and coded eye-event locations at specific AOIs (in Finnish).

the shared screen view, thereby creating a bi-directionally shared focus of attention (Gilbert 2007). Subsequently, Student B takes the lead (see DCRP analysis), experimenting with the lighting and requesting the partner to maintain a higher temperature. Although the process is not clearly verbalised, the coordination is evident through artefact manipulation and gaze patterns.

Time	Role	Event	Verbal	Eye gazing (AOI)
18.30	A	Move-resource	*Temperature*	3a
18.30	B	Move-resource	*Brightness*	3b, 4
18.30	A	Move-resource	*Temperature*	3b
18.30	A	Move-resource	*Temperature*	3b
18.30	B	Move-resource	*Brightness*	3b, 4
18.30	A	Move-resource	*Temperature*	3a, 4
18.30	B	Move-resource	*Brightness*	3b, 4, 3a
18.31	B	Move-resource	*Brightness*	3b, 4, 3a
18.31	A	Chat	*What about this other one then*	2, [3b, 5, 3b]
18.31	A	Chat	*Good*	2
18.31	A	Chat	*Now they are correct* *[referring to the position of temperature and lightning]*	2, 3a, 5, 2

Eye gazing (scan path view)

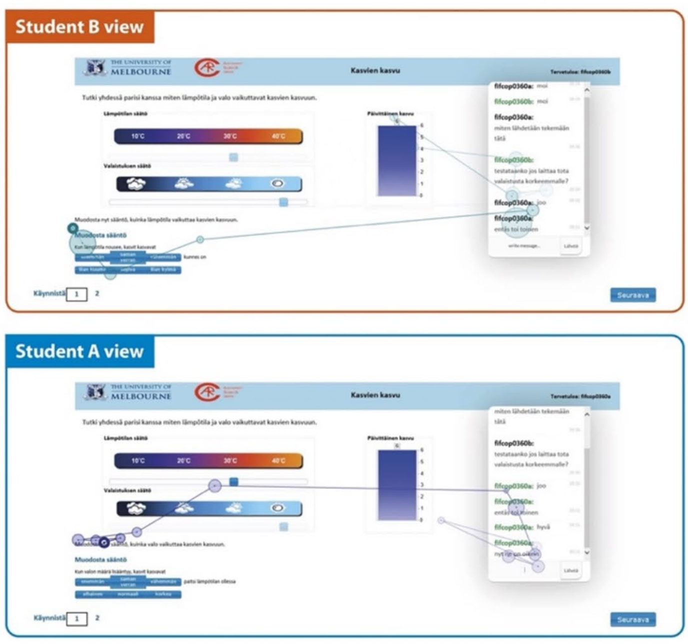


FIGURE 7 | Example from a shared attention-level experience (Dyad 8, Plant Growth). The episodes include logfile data, scan paths, and coded eye-event locations at specific AOIs (in Finnish).

Student A engages similarly, albeit with even less explicit verbalisation. In terms of the CPS construct, the verbal communication (i.e., interaction) observed in the dyad primarily involves negotiations (i.e., negotiate). Student B takes the initiative in progressing the task (i.e., responsibility initiative) by figuring out the rules and initiating discussions, while Student A contributes by enacting the rules through responsive actions (i.e., action).

3.2.3 | The Manifestation of Sliding Between Mutual and Shared Attention: Dyad 8

Based on their eye-gaze movements, both students begin Page 1 by carefully navigating the screen, exploring the manipulatives, and reading the instructions—yet without verbalising or sharing this information with one another. At the start of the task, they greet each other by saying, "Hi," thereby signifying knowledge of each other's presence (Gilbert 2007). Next, without sharing any detailed information of their screen views, they agree to begin testing, starting with lighting—an idea proposed by Student B. As in Taylor (1985), they put it out "between us" as something to focus on together (Siposova and Carpenter 2019, p. 262). Student B takes the lead in testing, which corresponds with the DCRP findings. Although no verbal exchange accompanies their testing, their eye movements suggest mutual tracking of each other's actions. After a while (see Figure 7), Student A verbalises the current situation (while Student B still silently tests the controls): "What about this other one then," followed by a reply, "Good," and finally "Now they are correct" (when the growth is the highest). However, when they define the rule, Student A—only at the end—realises that their rules differ, but this recognition is not jointly discussed. Throughout the task, their attentional experiences slide between mutual and shared, predominantly the latter. While they demonstrate a relatively good balance of communication (i.e., action and interaction), their exchanges lack intentionality. Neither partner clearly assumes responsibility for progression (i.e., responsibility initiative), checks their mistakes (i.e., self-evaluation), nor consults one another on how best to proceed (i.e., negotiating).

4 | Discussion

4.1 | Key Observations and Methodological Insights

This exploratory study investigated the social domain of CPS by focusing on JA behaviour as a sliding scale of jointness between two students interacting remotely, as defined by Siposova and Carpenter (2019). Overall, the findings underscore the value of JA for deepening understanding of the social domain of CPS, including in remote interaction contexts. Distinguishing between levels of JA was considered crucial for understanding the quality of social interaction during CPS, as reasoning about others from a third-person perspective differs significantly from directly engaging with others within a second-person perspective (Siposova and Carpenter 2019).

Methodologically, the study employed investigator, data, and method triangulation (Flick 2017), using multimodal data analysed through complementary quantitative and qualitative approaches. The strength of eye-gaze coupling served as a promising quantitative indicator of JA in a remote setting, guiding the selection of dyads for the qualitative analysis of interaction, revealing behaviours associated with higher attentional levels. The observed JA ranged from mutual to shared attention, involving intentional communication and sharing of attention. While the analysis of interaction yielded important insights into the verbal construction of JA, the incorporation of gaze data enabled the examination of its non-verbal dynamics (e.g., Whitehead et al. 2024). Thus, alongside the rich information embedded in the logfiles, gaze data offered additional insights into CPS processes, extending beyond what was observable in verbal interaction. In doing so, it rendered otherwise *invisible social moments visible* during CPS.

4.2 | Interpreting Social Domain of Remote CPS Through JA Behaviour

The qualities related to higher attentional levels closely align with attributes of "ideal" social behaviour in productive CPS, where participants are expected to acknowledge their partners, ask questions, and share information and resources (Scoular et al. 2017). In this study, the selected dyads displayed mutual awareness and coordination while working to solve the task. Instances of leadership emerged in each dyad, with one member taking more responsibility in planning and guiding progress, while the other remained largely responsive. However, as the CPS process progressed, reciprocal responsiveness became more prominent and visible, particularly through patterns of eye gazing.

Moreover, while attentional levels lie on a scale of jointness and may be nested within a single interaction episode, each level can also vary in strength (Siposova and Carpenter 2019). In the selected dyads, attentional level often slides between mutual and shared within a single task page. As observed in this study, lower attentional levels can form the foundation for higher ones. For example, mutual attention, such as acknowledging one s presence, may be crucial for achieving shared attention.

What, then, causes the rightward shifts along the scale of jointness? In this study, the co-participants' proximity, shared goals, similar tasks and clearly bounded online perceptual space may have increased confidence of attending together with the partner, enabling sliding along the scale of jointness (Siposova and Carpenter 2019). Although the remote setting, relying on chat and observable actions (e.g., adjusting temperature and lighting), is more limited than co-located CPS contexts, common ground may still emerge. While often described as a cognitive representation in the minds of the interlocutors (Clark and Brennan 1991), common ground, one of the factors that may cause shifts in the scale of jointness, can also arise through co-presence and inter-bodily coordination (Cowley and Harvey 2016), as observed in JA between a child and caregiver (Fowler and Hodges 2016). In remote contexts, where eye contact, bodily orientation, and pointing gestures are missing (Carpenter and Liebal 2012), even the "mere belief of being seen" (Cavallo et al. 2015, p. 67) may have an effect.

Additionally, the shift from individual attention to joint may occur through either top-down or bottom-up processes (León 2022). In top-down processes, one participant actively directs the partner's attention toward a joint target, whereas in bottom-up, stimulus-driven processes, a salient stimulus attracts their attention without goal-directed behaviour (Carpenter and Liebal 2012). When (verbally) acknowledged (Siposova and Carpenter 2019), through both processes, the participants can become aware of attending together to the same object (León 2022). However, in game-like environments, highly eye-catching stimuli may delay reading instructions or initiating contact with a partner, undermining the task's intended design.

Yet, the shift from solitary attention and first-person perspective to JA and second-person perspective is not always smooth (León 2022). Understanding others primarily arises through interaction rather than observation (Zahavi 2015). In remote settings, establishing shared understanding often requires additional efforts, potentially leading to attempts to achieve joint goals with limited sharing (Alterman and Harsch 2017). Likewise, status problems and imbalances of power frequently manifest in dominant interaction patterns (Langer-Osuna 2016). Status—understood as a relationship of power among dyads—may stem from academic skills, perceived intelligence, or social popularity (Langer-Osuna 2016). How power relations and intellectual authority are negotiated in interaction can, at worst, hinder student performance and dyadic productivity (Simpson et al. 2017).

Moreover, when a joint, second-person perspective is achieved, JA opens for the partners the possibility of "a shared action space in which the agents reframe and recalibrate a situation in light of the presence of a co-attender, in order to achieve joint goals" (Pezzulo et al. 2013, pp. 21, 22). This resonates well with the concept of shared problem space (Roschelle and Teasley 1995), the core operational concept pertaining to the cognitive aspects of CPS (e.g., task regulation), materialised by developing strategies for problem-solving and shared problem representation (Hesse et al. 2015). In this regard, recognising the different attentional levels and understanding what kinds of interactions and commitments they can support and what kinds of consequences they have on the interactional processes is directly applicable to designing and better supporting productive CPS.

4.3 | Limitations and Future Directions

This study is subject to several limitations. Due to the small number of dyads and the exploratory nature of the research, the aim was not to generalise the findings to broader populations. Instead, the study sought to develop a deeper understanding of the complex social domain of CPS through the lens of the theoretical construct of JA, by combining quantitative eye-tracking analysis with an in-depth qualitative approach. As Godfroid and Hui (2020) note, it is generally recommended to examine eye-tracking data alongside additional data sources and methods. While such data can indicate *where* individuals are looking at, they do not unequivocally explain *why* attention is directed in that way. This limitation highlights the importance of cautious interpretation and the value of triangulating gaze data with complementary qualitative or contextual information.

To deepen understanding of the processes underlying diverse attentional levels in CPS, further research is needed involving larger student populations including lower-performing students, varied contexts, and more advanced methodologies. In this study, CRQA was used exploratorily to assess whether gaze coupling occurred between partners (Wallot and Leonardi 2018). Building on this, future research could examine the temporal interplay between gaze and interaction sequences during CPS (Nüssli et al. 2013), which would require the analysis of longer interactions and behaviours surrounding shared attention episodes in more complex tasks. Although DCRP measures, signalling leader-follower behaviour (Dindar et al. 2020), showed diversity in social dynamics among the selected dyads, this was not the primary focus of the study. As no data were collected on individual characteristics such as personality traits (e.g., Stadler et al. 2019), the underlying causes of these asymmetries remain unexplored and warrant further investigations.

Additionally, computer vision (CV) approaches offer a promising avenue for enhancing the automation of AOIs and scalability of eye-tracking analysis, particularly by improving the estimation of gaze direction and fixation points, especially when combined with machine learning models to enhance understanding of visual attention (e.g., Chang et al. 2023; Shilaskar et al. 2023). However, automated AOI detection tends to be more applicable in studies involving dynamic stimuli or naturalistic viewing conditions. Furthermore, machine learning models would generally require substantially larger datasets for training and validation than were available in the present study. With appropriate dataset characteristics and scale, such methods offer a valuable direction for future research on JA.

5 | Conclusions

Research on JA behaviour holds significant potential for advancing our understanding of the complex social domain of CPS, even within remote interaction contexts (e.g., Pöysä-Tarhonen et al. 2021). This study contributes to this growing field by demonstrating the value of combining multiple methods and drawing on multimodal data, offering an integrated lens through which to examine how social interaction unfolds during CPS. However, as Baker and Reimann (2024) emphasise, working with data of high temporal granularity—such as eye-gaze measures—demands careful theoretical and analytical integration. Such data do not inherently reflect the situatedness of cognition and interaction but rather tend to abstract from context. In this study, the qualitative phase was essential in contextualising and enriching the interpretation of quantitative measures (Plano Clark 2019), enabling a more nuanced understanding of how JA emerges and functions within dyadic CPS tasks. Importantly, the findings highlight the value of distinguishing between different attentional levels as a means of evaluating the quality of CPS interactions. In doing so, the study lays important groundwork for future research into how task and learning environment design can be structured to support higher attentional levels and thereby promote more productive CPS processes and improved outcomes—supporting learners to "improve collectively" (Wise and Schwarz 2017, p. 446).

Author Contributions

Johanna Pöysä-Tarhonen: conceptualization, formal analysis, investigation, methodology, supervision, visualization, writing – original draft. **Shupin Li:** formal analysis, methodology, visualization, writing – original draft. **Jarkko Hautala:** data curation, investigation, methodology, supervision, writing – original draft. **Nafisa Awwal:** conceptualization, formal analysis, writing – original draft. **Päivi Häkkinen:** funding acquisition, writing – review and editing.

Acknowledgements

This study was funded by the Academy of Finland (grant number 273970). Open access publishing facilitated by Jyvaskylan yliopisto, as part of the Wiley - FinELib agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Endnotes

[1] Here, the term "dyadic" refers to the relationship between self and an object of attention, not pair-level activity as it does elsewhere in this article.

[2] The digital environment from the ATC21S project was developed during the Assessment and Teaching of 21st Century Skills (ATC21S) project at the University of Melbourne (e.g., Care et al. 2018; Griffin and Care 2015).

References

Alterman, R., and K. Harsch. 2017. "A More Reflective Form of Joint Problem Solving." *International Journal of Computer-Supported Collaborative Learning* 12: 9–33. https://doi.org/10.1007/s11412-017-9250-1.

Andrist, S., W. Collier, M. Gleicher, B. Mutlu, and D. Williamson Shaffer. 2015. "Look Together: Analyzing Gaze Coordination With Epistemic Network Analysis." *Frontiers in Psychology* 6: 1–15. https://doi.org/10.3389/fpsyg.2015.01016.

Andrist, S., A. R. Ruis, and D. Williamson Shaeffer. 2018. "A Network Analytic Approach to Gaze Coordination During a Collaborative Task." *Computers in Human Behavior* 89: 339–348. https://doi.org/10.1016/j.chb.2018.07.017.

Awwal, N., M. Alom, and E. Care. 2015. "Proceedings of the 9th European Conference on Game-Based Learning." In *Game Design for Learning to Solve Problems in a Collaborative Environment*. Academic Conferences and Publishing International Limited.

Baker, M., T. G. Hansen, R. Joiner, and D. Traum. 1999. "The Role of Grounding in Collaborative Learning Tasks." In *Collaborative Learning: Cognitive and Computational Approaches*, edited by P. Dillenbourg. Elsevier ScienceElsevier Science.

Baker, M. J., and P. Reimann. 2024. "CSCL has Come of Age." *International Journal of Computer-Supported Collaborative Learning* 19: 1–10. https://doi.org/10.1007/s11412-024-09419-2.

Barron, B. 2003. "When Smart Groups Fail." *Journal of the Learning Sciences* 12, no. 3: 307–359.

Care, E., P. Griffin, and M. Wilson, eds. 2018. *Assessment and Teaching of 21st Century Skills: Research and Applications*. Springer. https://doi.org/10.1007/978-3-319-65368-6.

Care, E., C. Scoular, and P. Griffin. 2016. "Assessment of Collaborative Problem-Solving in Education Environments." *Applied Measurement in*

Education 29, no. 4: 250–264. https://doi.org/10.1080/08957347.2016.1209204.

Carpenter, M., and K. Liebal. 2012. "Joint Attention, Communication and Knowing Together in Infancy." In *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, edited by A. Seemann, 159–182. MIT Press.

Cavallo, A., O. Lungu, C. Becchio, C. Ansuini, A. Rustichini, and L. Fadiga. 2015. "When Gaze Opens the Channel for Communication: Integrative Role of IFG and MPFC." *NeuroImage* 119: 63–69.

Chang, F., J. Zeng, Q. Liu, and S. Shan. 2023. "Proceedings of the 2023 Symposium on Eye Tracking Research and Applications." In *Gaze Pattern Recognition in Dyadic Communication*. Association for Computing Machinery.

Chi, M. T. H., R. Glaser, and E. Rees. 1982. "Expertise in Problem Solving." In *Advances in the Psychology of Human Intelligence*. Lawrence Erlbaum.

Clark, H. H., and S. E. Brennan. 1991. "Grounding in Communication." In *Perspectives on Socially Shared Cognition*, edited by L. B. Resnick, J. M. Levine, and S. D. Teasley, 127–149. American Psychological Association.

Coco, M. I., and R. Dale. 2014. "Cross-Recurrence Quantification Analysis of Categorical and Continuous Time Series: An R Package." *Frontiers in Psychology* 5: 510.

Coco, M. I., D. Mønster, G. Leonardi, R. Dale, and S. Wallot. 2020. "Unidimensional and Multidimensional Methods for Recurrence Quantification Analysis With CRQA." *Geriatrics* 13: 145.

Cowley, S. J., and M. I. Harvey. 2016. "The Illusion of Common Ground." *New Ideas in Psychology* 42: 56–63. https://doi.org/10.1016/j.newideapsych.2015.07.004.

Dindar, M., S. Järvelä, S. Ahola, X. Huang, and G. Zhao. 2020. "Leaders and Followers Identified by Emotional Mimicry During Collaborative Learning: A Facial Expression Recognition Study on Emotional Valence." *IEEE Transactions on Affective Computing* 13, no. 3: 1390–1400. https://doi.org/10.1109/TAFFC.2020.3003243.

Flick, U. 2017. "Mantras and Myths: The Disenchantment of Mixed-Methods Research and Revisiting Triangulation as a Perspective." *Qualitative Inquiry* 23, no. 1: 46–57. https://doi.org/10.1177/1077800416655827.

Fowler, C. A., and B. Hodges. 2016. "Finding Common Ground: Alternatives to Code Models for Language Use." *New Ideas in Psychology* 42: 1–6. https://doi.org/10.1016/j.newideapsych.2016.03.001.

Funke, J., A. Fischer, and D. V. Holt. 2018. "Competencies for Complexity: Problem-Solving in the Twenty-First Century." In *Assessment and Teaching of 21st Century Skills: Research and Applications*, edited by E. Care, P. Griffin, and M. Wilson, 41–53. Springer.

Gilbert, M. 2007. "Mutual Recognition, Common Knowledge, and Joint Attention." In *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz,* edited by T. Ronnow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson, 1–21. Lund University.

Godfroid, A., and B. Hui. 2020. "Five Common Pitfalls in Eye-Tracking Research." *Second Language Research* 36, no. 3: 277–305. https://doi.org/10.1177/0267658320921218.

Gómez, J. C. 1996. "Ostensive Behavior in Great Apes: The Role of Eye Contact." In *Reaching Into Thought: The Minds of the Great Apes*, edited by A. E. Russon, K. A. Bard, and S. T. Parker, 131–151. Cambridge University Press.

Graesser, A., S. Greiff, M. Stadler, and T. Shubeck. 2020. "Collaboration in the 21st Century: The Theory, Assessment, and Teaching of Collaborative Problem Solving." *Computers in Human Behavior* 104. https://doi.org/10.1016/j.chb.2019.09.010.

Griffin, P., and N. Awwal. 2018. "A Process Perspective of Collaborative Problem Solving." In *International Trends in Educational Assessment*, 62–85. Brill.

Griffin, P., and E. Care. 2015. "The ATC21s Method." In *Assessment and Teaching of 21st Century Skills: Methods and Approach*, edited by P. Griffin and E. Care, 1–33. Springer.

Hesse, F. M., E. Care, J. Buder, K. Sassenberg, and P. Griffin. 2015. "Framework for Teachable Collaborative Problem-Solving Skills." In *Assessment and Teaching of 21st Century Skills: Methods and Approach*, edited by P. Griffin and E. Care, 37–56. Springer.

Just, M. A., and P. A. Carpenter. 1980. "A Theory of Reading: From Eye Fixations to Comprehension." *Psychological Review* 87, no. 4: 329–354. https://doi.org/10.1037/0033-295X.87.4.329.

Langer-Osuna, J. M. 2016. "The Social Construction of Authority Among Peers and Its Implications for Collaborative Mathematics Problem Solving." *Mathematical Thinking and Learning* 18, no. 2: 107–124. https://doi.org/10.1080/10986065.2016.1148529.

León, F. 2022. "Attention in Joint Attention: From Selection to Prioritization." In *Access and Mediation: Transdisciplinary Perspectives on Attention*, edited by M. Wehrle, D. D'Angelo, and E. Solomonova, 65–90. De Gruyter.

Marwan, N., M. C. Romano, M. Thiel, and J. Kurths. 2007. "Recurrence Plots for the Analysis of Complex Systems." *Physics Reports* 438, no. 5–6: 237–329. https://doi.org/10.1016/j.physrep.2006.11.001.

Moll, H., and A. N. Meltzoff. 2012. "Joint Attention as the Fundamental Basis of Understanding Perspectives." In *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, edited by J. Seemann, 392–413. Oxford University Press.

Moore, C., and J. Barresi. 2017. "The Role of Second-Person Information in the Development of Social Understanding." *Frontiers in Psychology* 8, no. 1667: 1–15. https://doi.org/10.3389/fpsyg.2017.01667.

Mundy, P. 2018. "A Review of Joint Attention and Social-Cognitive Brain Systems in Typical Development and Autism Spectrum Disorder." *European Journal of Neuroscience* 47, no. 6: 497–514. https://doi.org/10.1111/ejn.13720.

Nüssli, M. A., P. Jermann, M. Sangin, and P. Dillenbourg. 2013. "REGARD: Remote Gaze-Aware Reference Detector." In *Eye Gaze in Intelligent User Interfaces*, edited by Y. Nakano, C. Conati, and T. Bader, 63–84. Springer.

Olsen, J. K., V. Aleven, and N. Rummel. 2017. "Exploring Dual Eye Tracking as a Tool to Assess Collaboration." In *Innovative Assessment of Collaboration: Methodology of Educational Measurement and Assessment*, edited by A. von Davier, M. Zhu, and P. Kyllonen, 157–172. Springer. https://doi.org/10.1007/978-3-319-33261-1_10.

O'Madagain, C., and M. Tomasello. 2019. "Joint Attention to Mental Content and the Social Origin of Reasoning." *Synthese* 198, no. 5: 4057–4078. https://doi.org/10.1007/s11229-019-02327-1.

Pezzulo, G., P. Iodice, S. Ferraina, and K. Kessler. 2013. "Shared Action Spaces: A Basis Function Framework for Social Re-Calibration of Sensorimotor Representations Supporting Joint Action." *Frontiers in Human Neuroscience* 7: 800.

Plano Clark, V. L. 2019. "Meaningful Integration Within Mixed Methods Studies: Identifying Why, What, When, and How." *Contemporary Educational Psychology* 57: 106–111. https://doi.org/10.1016/j.cedpsych.2019.01.007.

Pöysä-Tarhonen, J., N. Awwal, P. Häkkinen, and S. 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, no. 1: 11. https://doi.org/10.1186/s41039-021-00160-0.

R Core Team. 2021. "R: A Language and Environment for Statistical Computing."

Roschelle, J., and S. D. Teasley. 1995. "The Construction of Shared Knowledge in Collaborative Problem Solving."

Schneider, B., and R. Pea. 2013. "Real-Time Mutual Gaze Perception Enhances Collaborative Learning and Collaboration Quality." *International Journal of Computer-Supported Collaborative Learning* 8, no. 4: 375–397.

Schneider, B., and R. Pea. 2014. "Toward Collaboration Sensing." *International Journal of Computer-Supported Collaborative Learning* 9, no. 4: 371–395.

Schneider, B., K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. 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: 241–261. https://doi.org/10.1007/s11412-018-9281-2.

Scoular, C., E. Care, and F. Hesse. 2017. "Designs for Operationalizing Collaborative Problem-Solving for Automated Assessment." *Journal of Educational Measurement* 54, no. 1: 12–35.

Sears, D. A., and J. M. Reagin. 2013. "Individual Versus Collaborative Problem Solving: Divergent Outcomes Depending on Task Complexity." *Instructional Science* 41: 1153–1172.

Shilaskar, S., S. Bhatlawande, T. Gadad, S. Ghulaxe, and R. Gaikwad. 2023. "Proceedings of the 2023 7th International Conference on Computing Methodologies and Communication (ICCMC)." In *Student Eye Gaze Tracking and Attention Analysis System Using Computer Vision*, 889–895. IEEE. https://doi.org/10.1109/ICCMC56507.2023.10083874.

Simpson, A., N. Bannister, and G. Matthews. 2017. "Cracking Her Codes: Understanding Shared Technology Resources as Positioning Artifacts for Power and Status." *International Journal of Computer-Supported Collaborative Learning* 12, no. 3: 221–249. https://doi.org/10.1007/s11412-017-9261-y.

Siposova, B., and M. Carpenter. 2019. "A New Look at Joint Attention and Common Knowledge." *Cognition* 89: 260–274. https://doi.org/10.1016/j.cognition.2019.03.019.

Stadler, M., K. Herborn, M. Mustafić, and S. Greiff. 2019. "Computer-Based Collaborative Problem Solving in PISA 2015 and the Role of Personality." *Journal of Intelligence* 7, no. 3: 15. https://doi.org/10.3390/jintelligence7030015.

Taylor, C. 1985. *Human Agency and Language*. Cambridge University Press.

Tomasello, M. 1995. "Joint Attention: Its Origins and Role in Development." In *Joint Attention as a Social Cognition*, edited by C. Moore and P. J. Dunham. Lawrence Erlbaum.

Villamor, M. M., and M. Rodrigo. 2019. "Gaze Collaboration Patterns of Successful and Unsuccessful Programming Pairs Using Cross-Recurrence Quantification Analysis." *Research and Practice in Technology Enhanced Learning* 14: 25. https://doi.org/10.1186/s41039-019-0118-z.

Vista, A., N. Awwal, and E. Care. 2016. "Sequential Actions as Markers of Behavioural and Cognitive Processes: Extracting Empirical Pathways From Data Streams of Complex Tasks." *Computers & Education* 92: 15–36.

Wallot, S., and G. 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: 2232. https://doi.org/10.3389/fpsyg.2018.02232.

Whitehead, R., A. Nguyen, and S. Järvelä. 2024. "Exploring the Role of Gaze Behaviour in Socially Shared Regulation of Collaborative Learning in a Group Task." *Journal of Computer Assisted Learning* 40, no. 5: 2226–2247. https://doi.org/10.1111/jcal.13022.

Wilby, M. 2023. "The Form and Function of Joint Attention Within Joint Action." *Philosophical Psychology* 36, no. 1: 134–161. https://doi.org/10.1080/09515089.2022.2039384.

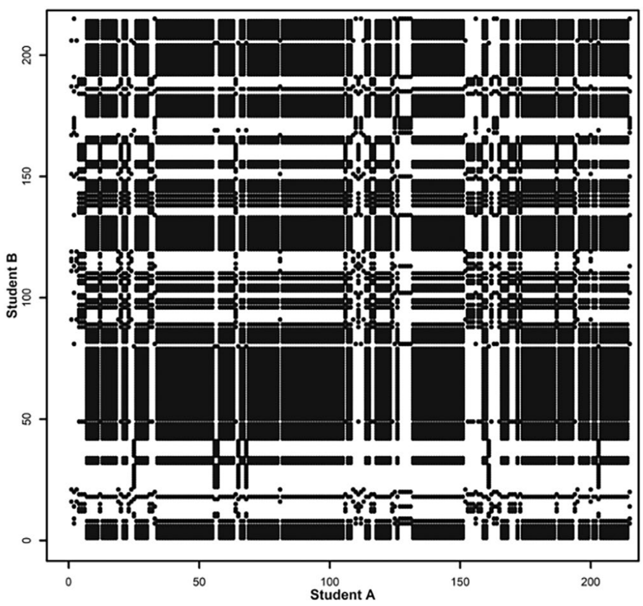
Wise, A. F., and B. B. Schwarz. 2017. "Visions of CSCL: Eight Provocations for the Future of the Field." *International Journal of Computer-Supported Collaborative Learning* 12, no. 4: 423–467. https://doi.org/10.1007/s11412-017-9267-5.

Wisiecka, K., Y. Konishi, K. Krzysztof, et al. 2023. "Supporting Complex Decision-Making: Evidence From an Eye Tracking Study on In-Person and Remote Collaboration." *ACM Transactions on Computer-Human Interaction* 30, no. 5: 1–27. https://doi.org/10.1145/358178.

Zahavi, D. 2015. "Self and Other: From Pure Ego to Co-Constituted We." *Continental Philosophy Review* 48, no. 2: 143–160. https://doi.org/10.1007/s11007-015-9328-2.

Appendix 1

Example of a cross recurrence plot (CRP).

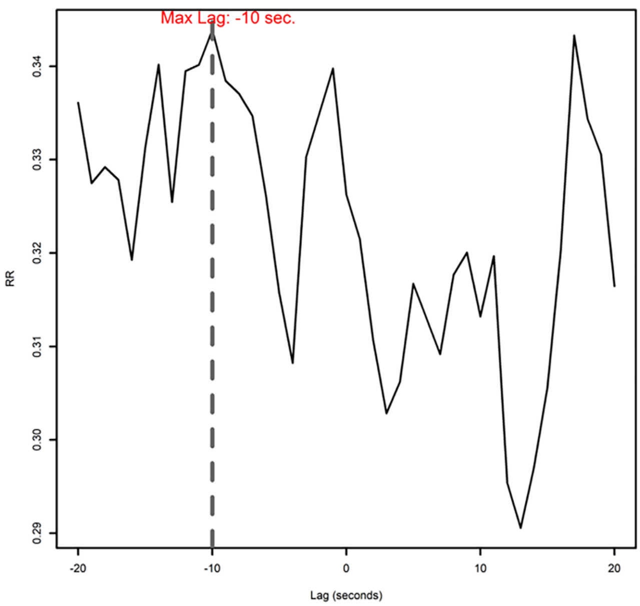


Note. The longest diagonal of the plot represents the gaze alignment at a lag of 0 (the time scale in seconds).

If the fixations are recurrent, they are shown as black points or pixels on the point. In this study, the labels along the horizontal and vertical axes refer to the fixation timelines of the first and second collaborators (Students A and B, respectively). Diagonal slices (lower-left to upper-right) correspond to an alignment of the gazes within a particular time lag between them, and a point is plotted on the diagonal when the gaze is recurrent. Alongside, varied small-scale structures (i.e., textures) can be seen, such as single and isolated recurrence points, reflecting random or strong fluctuations in the data, or bands of white space, indicating i.e., that the collaborators looked at different spots on the screen (Marwan et al. 2007).

Appendix 2

Example of a diagonal cross recurrence profiles (DCRP).



Note. The percentage of recurrence points is plotted on the y-axis, the lags around LoS on the y-axis. In the example, the two time series are most strongly coupled at the lag of −10 where a peak in the DCRP can be observed (see the dashed line).