Information flow and children's emotions during collaborative coding: A causal analysis

Kshitij Sharma
kshitij.sharma@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

Sofia Papavlasopoulou
spapav@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

Serena Lee-Cultura
serena.leecultura@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

Michail Giannakos
michailg@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

ABSTRACT

This paper investigates the relation between children's joint gaze and emotions with the information flow of the screen from a causal point of view, in the context of collaborative coding. We employ Granger's definition of causality to extend the knowledge we have about children's collaborative activities from correlational methods. We organised a coding workshop with 50 children (10 dyads and 10 triads; 13-16 years old). While the children were coding collaboratively, their facial video and the screen were recorded. From the screen recording we computed the information flow; and from the facial video we computed children's emotions (e.g., frustration and boredom) and estimated their gaze. The gaze estimation was used to compute the joint visual attention (JVA) of the team. Our results show that for high performing teams JVA drives the information flow; while for low performing teams we observe causal relation between emotions and information flow. In particular for the low performing teams, frustration and boredom drive the information flow and the information flow then drives children's confusion. These results extend the understanding of the socio-cognitive processes underlying collaborative performance, which is primarily correlational in nature, with the causal relations between measurements. These novel results have the potential to guide the design of learning tools that scaffold children's learning and collaboration.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**.

KEYWORDS

Collaborative learning, Computational thinking, Joint visual attention, emotions, affective states, causal modelling, Granger causality

IDC '21, June 24–30, 2021, Athens, Greece
Published paper can be found in here:
https://doi.org/10.1145/3459990.3460731

ACM Reference Format:
Kshitij Sharma, Sofia Papavlasopoulou, Serena Lee-Cultura, and Michail Giannakos. 2021. Information flow and children's emotions during collaborative coding: A causal analysis. In *Interaction Design and Children (IDC '21), June 24–30, 2021, Athens, Greece.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3459990.3460731

1 INTRODUCTION & MOTIVATION

As the phenomena surrounding the interaction between children and computational and communication technologies advance [37], so does the data produced from this interaction. During the last years, researchers leverage on such data to portray children's interaction trajectories/paths and support the design and development of both the digital technology and the respective processes (e.g., instruction, treatment). In the context of collaborative learning, data collected from child-computer interaction (CCI) are utilized to support children's cognitive and learning processes (e.g., adaptive interfaces) [33, 68] as well as the quality of collaboration [98]. It is important to support children's collaboration due to the fact that the collaborative learning is deemed as one of the 21st century skill [105] not only for the adults [113] but also for the primary and secondary education [11]. Collaborative learning is also important for developing critical thinking capabilities [60] as well. To regulate children's expectations and improve collaboration, we have seen in the literature the use of adaptive collaborative learning support (ACLS) [43, 48, 81, 115]. ACLS leverages on intelligent technologies to improve children collaboration and learning by identifying the current state of the interaction and providing a tailored intervention. For such systems to effectively operate, it is important to be able to use highly temporal data to identify children's interaction and communication at the micro level.

To support the design and development of the aforementioned technologies, this paper proposes the use of gaze-tracking (from the frontal video and not eye-tracking), and facial-expressions (as an input from the child) and the information representations of the screen (as an input from the system) to assess children's interaction and collaboration, through the lens of the causal relationships during CCI. We study these causal relationships while the children are engaged in a collaborative coding workshop in dyads and triads. Specifically, we investigate two causal relationships. First, between children's joint visual attention (i.e., how much they look at the same area at the same time) and the information representations of the screen. Second, between children's joint emotional state (i.e., how similar their emotions such as, boredom, confusion and frustration are, at the given time) and the information representations of the screen. For this analysis, we used time series data from the children working on the Scratch code. We further investigate the nature of the causal relationships among aforementioned measurements with respect to children's performance levels (high/low) in the collaborative coding task.

Why causality is important? A causal relation between time series of two measurements represent an essential "active connection" between them [35, 116]. Untangling correlation and causation has recently received increased attention across multiple disciplines, with documented benefits in both HCI [82, 112] and learning [94, 96]. Investigating the causal relationship allows us to understand deeper the causal interactions between children and technology, ultimately leading to the design of improved interactive technology. Causation goes far beyond correlation between variables and accounts for the "information exchange" (interaction) and the underlying processes that are responsible for the two variables under investigation. For example, there is a vast collection of studies that show the correlation between the information content and the joint visual attention in collaborative scenarios [7, 85, 97]. However, in absence of a causal link between the two measurements (i.e., joint visual attention and information content) it becomes difficult to design real-time adaptive systems to support processes that enable efficient consumption of the information [69]. Causal relations allow us to account for relationships that are inextricably intertwined with humans' interactions (e.g., emotions, see [82] and provide the decision makers (e.g., designers, researchers) a stronger basis (as compared to correlations) to decide upon the necessary actions for a given desired result [69]. This paper exemplifies how causal analysis can support CCI research by showing the direction of causality between screen's information representations and children's joint visual attention, and screen's information representations and children's joint emotional states, in the context of collaborative coding workshop. We also show how these relationships change based on the task-based performance levels (high and low) of the children. Specifically, we address the following research questions:

RQ1 What is the causal relation between information flow and joint visual attention of children during coding?

RQ2 What is the causal relation between information flow and joint emotional state of children during coding?

RQ3 How are these causal relations affected from children's performance levels?

To tackle the aforementioned research questions, we conduct a study in which we use frontal videos to capture children's gaze and emotions during a coding activity. The data is coupled with the information presentation (captured with screen recording) and children task-based performance (captured with the produced artifacts). We then apply Granger causality to investigate the causal interactions between children and technology during collaborative coding. By investigating the casual interactions we provide a path towards a technology design that support children's collaboration and learning. In particular, this works contributes to CCI in three ways: **Methodologically,** utilizing children's facial videos and screen recordings to study the relationship between their collaboration and emotional state with the information presented. **- Analytically,** showing that children's emotions drive the use of block-based programming technology (e.g., Scratch) more often than the other way around, and providing several insights on this relationship. **- Conceptually,** highlighting implications for practice and research in both technology design and facilitation to support collaborative coding activities for children.

2 RELATED WORK

2.1 Joint Visual Attention in education and CCI

In collaborative learning and problem solving contexts, Joint Visual Attention (JVA) has been used to explain the socio-cognitive processes underlying expertise [50], performance [92] and learning gains[73]. JVA has been computed as the cross-recurrence (discrete) of the gaze-behaviour of the collaborators [34, 78], and as the similarity of two gaze-patterns (continuous) [73, 92]. In both cases, JVA is a measurement of collaborators looking at the same set of objects or similar parts of the stimulus at the same time or in the same time window.

Specifically, students having high JVA during their collaboration translates to a higher level of "mutual understanding" of the concept used [91]. Collaboration involves multiple verbal and deictic references to the visual task at hand [49]. In that way, the peers establish a "common ground" to work/learn together[17, 18]. By using eye-tracking, it is shown that peers who follow each other's deictic [49] and verbal [34] references perform better than those who do not. In addition, early eye-tracking studies show that there is a coupling between speaker's gaze and the listeners' gazes that tells us about how well they are following each other in the collaborative space [1, 41, 78]. Another phenomenon that explains the high collaborative performance is the "convergent conceptual change" (CCC) [80, 93, 108]. When peers collaborate, they change the conceptual understanding of the learning material/domain [80]. It is shown that teams who collaborate better, have less difference between peers' understanding after, than that before the collaboration [13, 80, 108]. In other words, high performing peers have a converging change in their conceptual understanding, while this can not be said about the low performing pairs [13, 80]. In recent eye-tracking studies, it was shown that JVA can explain CCC among peers in collaborative learning settings [93].

In CCI research, JVA has been used to explain/predict the coding performance [73], learning gain [66] and expertise [97]. In a collaborative coding study, it was shown that children with high JVA (gaze similarity) have higher learning gains than those with low JVA [73]. Various studies, in different contexts, have shown that JVA can explain the collaboration quality of the teams (e.g., learning to code [73, 97]; fractions [7, 66]; neuroscience [84]). Furthermore, Sharma et al. showed that the expertise moderates the relation between JVA and learning outcome [97]. In all the aforementioned studies, the relation between the JVA and collaborative processes has been examined from a correlational view-point. The aim of this study is to analyse the causality between the children's JVA and the screen's information flow and show how the causal relation changes based on children's collaborative coding performance.

2.2 Emotions in education and CCI

CCI researchers, have measured emotions via various modes of data collection, for example, self reports [36, 89], physiological data [23, 57, 101, 107], and facial features [53, 56, 110]. Children's emotional/affective state is an important topic in CCI, with several studies focusing on usability [36, 110], learning processes [101], performance [98], , engagement [57] and enjoyment [56]. Regarding the emotions that are considered important in educational settings, there are two primary strands. The first one is motivated by the Control Value Theory (e.g., happiness, sadness, anger, surprise, disgust) [74] and the second one uses the affective states (e.g., frustration, confusion, delight, boredom) [29]. Previous work has shown that happiness is related to success [32] and anger to failure [8]. Furthermore, frustration appears as a common feeling among students involved in online collaborative learning experiences [61]; whereas, boredom and confusion are related to poor academic performance [5, 28]. Further, emotions/expressions/affective states have been used in educational research to provide feedback [100, 110, 117], improve students' interaction [44, 75, 106] and performance [51]. For a detailed review affective states and emotions in education please see [77].

While students collaborate in front of a computer during a coding task, they are socially engaged with the same goal to successfully create an artifact [72]. During students' collaboration, it is important to maintain durable relationships and acceptable levels of participation. Interactions that are associated with these aspects of the group performance can be typified as social-emotional interactions [55]. In collaborative settings, frustration was found to be prominent during online interaction [10] and online discussion forums [14]. Confusion occurs when the groups have to reinforce their pre-existing mental models with new information [16, 29], and was also found to lead to impasses in collaborative learning [114]. Finally, boredom is mostly observed in the cases where the problem at hand is far too easy or repetitive[70]; as it is the case with individual learning [19]. In this paper, we decided to focus on these three emotions because these emotions were found to be most prominent in a selective meta-study with 21 studies [26].

Regarding Joint Emotional State (JES) during collaborative learning, in a recent study [98], it was shown that the JES was correlated with the perceived collaborative performance. As it is the case with JVA, all the aforementioned studies present the relation between the JES and collaborative processes from a correlational view-point. Additionally, there are only few studies that analyse children's emotional states (or facial expressions) using a collaborative measure. At the same time there is a need for further research to examine children's collaborative coding processes using the rich nonverbal communications such us facial expressions [111]. Most of the studies use individual emotions or the average value of emotions as the collaborative measure. In this paper, we will analyse the causality between the children's JES (a collaborative measurement of their emotional states) and the screen's information content (information flow) and show how the causal relation changes based on children's collaborative coding performance.

2.3 Exploring Causality in HCI and education

Causal analysis is not new in the fields of neuroscience [25, 38], economics [46, 52], and life-sciences [79]. However in HCI and education, formalising causal relations involving multiple individuals and groups is a rather recent practice [54, 119]. For example, Van Berkel and colleagues showed that the number of notifications on a mobile phone causes the number of times the user switches on the screen [112]. Further, Sarsenbayeva and colleagues, showed that the emotions such as, contempt, disgust and joy cause application launch events such as, communication, social and work [82]. On the other hand, sadness and surprise are cause by launching social and communication applications [82].

In digital education, it was shown that for instruction style had a mediation effect on the direction of causality between the students' cognitive load and information flow [94]. For static instruction (learning code-debugging) information flow causes cognitive load, while for dynamic instruction (video-based learning) cognitive load causes information flow [94]. In collaborative learning situations involving children, recent research have focused on explaining the causal relation between individual gaze and JVA [95, 96]. In two different studies, first it was shown that for high performing students JVA was causing the individual gaze; and for the low performing students the individual gaze was causing the JVA [95]. Second study showed that when high performing students were involved in problem solving dialogue their JVA was causing the individual gaze [96]. The aforementioned results are in individual settings [82, 94, 112], discrete events [82, 112], or investigating causal relations between uni-modal measurements [95, 96] (e.g., gaze-based variables). In this paper, we present the causal relations between multi-modal measurements (e.g., gaze, emotions, screen's content).

3 METHODS

3.1 The coding activity

We designed and implemented a coding activity at the Norwegian University of Science and Technology (NTNU), Trondheim, Norway. The activity follows the constructionist approach, and the main principles of "Making" [71]. The workshop had an informal environment and was organised, as an out-of-school activity. Children ranging from 13–16 years old, were invited to participate in the activity, which did not require any previous knowledge of coding from them. Specifically, children were introduced to block-based programming through Scratch and were instructed to modify and develop their own games, working collaboratively in dyads or triads (depending on the number of children). Student assistants were the instructors of the activities and were supporting children's teams as needed. Each instructor was observing one or two teams. In addition, three researchers were also present in the implementation of the workshops and apart from taking care of the execution of the workshop, they were observing and writing notes. The workshop was divided into two sections and lasted for approximately 4 hours. Children created their games step by step by iterative coding and testing them. After completing the games, all teams reflected and played each others games. The session lasted approximately three hours.

3.2 Sampling and data collection

The study was conducted in Autumn 2017; children from 8^{th} to 10^{th} grade (age 13-16 years old) participated in the coding workshop, after their school–teacher applied to attend it. The sample consisted of 105 participants in total, 69 boys and 36 girls (mean age: 14.55, SD: 0.650). We video recorded 10 dyads and 10 triads while they were coding their games. The children were randomly assigned to their dyads or triads. Following all recommendations and regulations for research which children in Norway, data were collected after having the necessary consent from both the child and the legal guardian. The invitation for participation was sent to almost all the schools in the Trondheim region with children in that age that had the possibility to bring them in our premises; from those, all schools who responded and showed interest were accepted to participate in our workshops. In particular we used:

Video recording: To capture children's facial expressions while coding their game, and detect their emotions, we used a wide-angle Logitech Webcams. The web camera was zoomed at 150% into the children's faces capturing video at 10 FPS. In total, we collected videos from 50 children (29 females), 10 triads and 10 dyads.

Artifacts (developed games): We collected the games as artifacts created from of the children. For each team, four game versions were saved. The first version was saved 45 minutes after they started coding and since then, each game version was saved every 45 minutes. This decision was suggested from the instructors who run the coding workshops and have a lot of experience with how the children in the specific workshop are progressing with the activity they need to accomplish. According to them, saving the game every 45 minutes is a good time frame that allows us to monitor the children's progress (e.g. not too short or too disturbing, not too long to lose monitoring important part of progression).

3.3 Measurements

3.3.1 Joint Visual Attention (JVA). This is defined as the proportion of time that children of the same team spend looking at the similar set of objects in the same given time window. Following are the steps to compute the JVA from the facial video of the collaborating children: *a)* Detect the faces in every frame of the video. *b)* Compute the gaze direction from the facial image in the frame for each face detected (by giving a 3D vector pointing at the screen, see Figure 1 top-left). OpenFace provides a 3D vector towards the center of the camera to represent the gaze direction. *c)* Extend the 3D vector of the gaze direction to intersect the plane of the computer screen. The point of intersection of the 3D gaze-vector and the plane of the computer screen provides the point where the child is looking at. We call this gaze-point (Figure 1 top-left). *d)* Once we have the gaze point of all the participants in the frame, divide the whole screen space into a grid of 20 rectangles (four rows and five columns) and assign the gaze-point to one of the 20 rectangles (Figure 1 top-right). *e)* Compute the cross-recurrence [78] between all the children present in the video to identify their JVA; and normalise using the group size (2 or 3).

3.3.2 Joint Emotional State (JES). This is defined as the proportion of time that children spend in a given emotional state (i.e., frustration, boredom, confusion) in the same given time window. We selected these states because these were found to be the most prominent across a wide range of studies [5, 26, 27]. Following are the steps to compute the JES from the facial video of the collaborating children: *a)* Detect the faces in every frame of the video (Figure 2 top-left). *b)* Align the faces across the frames so that same faces are being tracked and assigned the same ID in every frame by using the method described in Sharma et. al. [98] (Figure 2 bottom-left).

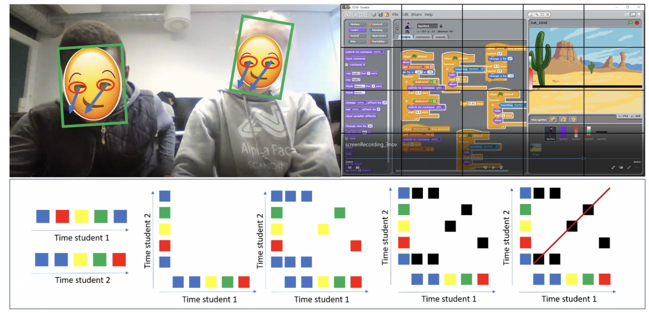

Figure 1: Top-left: example of multiple face detection, with the gaze direction estimation. Top-right: example of the grid layover on the stimulus (code on the screen). Bottom: Summary of steps from the individual time series of gaze on the boxes to group's cross recurrence (JVA). Each colour represents a different box on the screen.

c) Once we have the faces with correct IDs, use OpenFace [2, 6] to compute the Action Units (AUs) [42] for each frame (Figure 2 right). *d)* From the AUs compute the proportions of the three emotions: frustration, boredom and confusion [62] during a fixed time window of ten seconds [98]. We used a generalised additive model to combine the AUs to compute the expressions [59]. Frustration was computed as a combination of AU12 and AU43; boredom was computed as a combination of AU4, AU7, and AU12; finally, confusion was computed as a combination of AU1, AU4, AU7, and AU12. *e)* Once we have the proportions of the three emotions, compute the cosine similarity among the probabilities for each children.

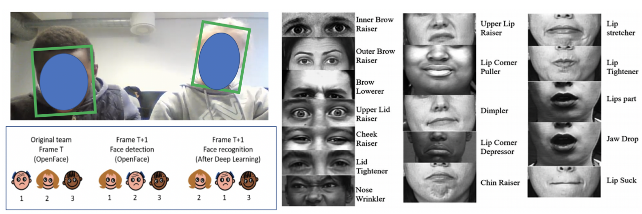

Figure 2: Top-left: example of multiple face detection. Bottom-left: example of providing a consistent ID to all the faces with face tracking using facial recognition. Right: Set of typical Action units.

3.3.3 Information Flow. Information flow (i.e., stimulus entropy) was computed for each frame of the screen recording. This indicates, in a direct manner, the amount of information transmitted to the student via the screen. We would like to point out here that our definition of information flow is slightly different from the definitions used in the information theory research [58, 99]. Smith (2009) uses the amount of information exchange between two nodes (input and output) as the definition of information flow, which is prone to loss of information [99], while Liang (2018) provides multiple definitions of information flow based on the Shanon entropy (similar to our definition) and divergence based methods [58]. In our case, we depend on the Shannon entropy of one node/channel only.

To compute the stimulus entropy for each frame, using a window of 10 seconds, we have used the three separate grey images (one each for red, green, and blue channels). This gives us three 2D arrays of values between 0 and 255[1]. We then compute the Shannon entropy of these three arrays using the following formula. This is a widely used method to compute RGB image entropy in image processing applications [39]. The mean entropy of the three arrays gives us the stimulus entropy (i.e., Information Flow). As indicated before, Shannon entropy is a direct measurement of the information content of the communication medium. This can be seen in the different cases shown in the Figure 3 for the coding workshop. We can see how the amount of information present on the screen changes with content of the screen. The entropy values increase from top to bottom panels, as it can be also seen from their color histograms, which show the information content on the screen.

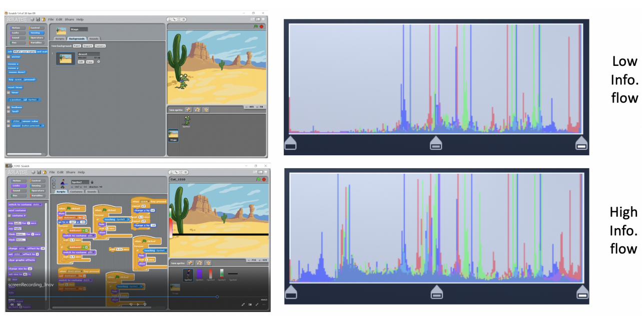

Figure 3: Typical examples for stimulus information flow calculation. We can visually notice that the top figure has low amount of information and the bottom figure has high amount of information.

The mean entropy of the three arrays gives us the stimulus entropy (i.e., Information Flow). As indicated before, Shannon entropy is a direct measurement of the information content of the communication medium. This can be seen in the different cases shown in the Figure aa for the coding workshop. We can see how the amount of information present on the screen changes with the content of the screen. The entropy values increase from top to bottom panels, as it can be also seen from their color histograms, which show the information content on the screen.

3.3.4 Coding Performance. Children's coding performance is based on the artifacts (Scratch code) that has been collected every 45 minutes. Those artifacts were used as the basis for computing their progress based on a widely used tool, called DrScratch [64]. DrScratch provides a fine-grained analysis of Scratch projects by supporting the assessment of computational thinking (CT) skills, using seven CT components: parallelism, logic, flow control, data representation, abstraction, user interactivity, and synchronisation. DrScratch allows the automatic, easy and quick analysis of Scratch projects providing feedback based on the score [64] with the results indicating comparable assessment with the one of a human expert [65]. Previous research has also used DrScratch to look at the development of each CT component as students design their games over time [109].

Each project (in our case all four versions of the games created from the teams) was uploaded and analysed by DrScratch system. The results give a general score to the project (i.e., max 21) which is the sum of the individual scores for each of the seven CT components (i.e., from 1 to 3). After having the general scores for each version of the games, we computed the gain between two consecutive versions. In that way, we have three "performance gains" for each team. For the rest of this paper, we will refer to "coding performance gains" as "gains". Then, based on these the three gains, we use a median cut to divide the teams of children into high and low gain groups. This happened for all the three different gains (i.e. from the first until the forth and last version of the game) corresponding to the three phases.

3.4 Data Analysis: Granger causality for more than one pair of time series data

To identify the casual relations between the measurements, we employed Granger causality [40] test. Granger causality has two assumptions [40], first is that cause occurs before effect and second is that the cause has information about the effect that is more important than the history of the effect. In terms of the nature of the concerned time-series, Granger causality is defined for linear and stationary time-series contexts, but variations for non-linear [3, 24] and non-stationary [15, 45] contexts exist. Granger causality is one of the various data analysis techniques to identify causal relations (e.g., Convergent Cross Mapping is also used in HCI studies [82, 112]). Granger causality was selected since it has proven usefulness in the context of learning technologies [94, 96]. For more details and the mathematical formulation of Granger Causality in the context of learning technologies please see [94, 96]. Here we provide a summary of steps so the method can be replicated based on this paper only. Lets assume that we are modelling the Granger causality between two variables X and Y.

(1) For each group compute the following:
   (a) Compute the partial η^2 for the model "X Granger causes Y".
   (b) Compute the partial η^2 for the model "Y Granger causes X".
   (c) Compute the partial η^2 for the model "Y linearly predicts X". This is similar to a correlational model.
   (d) Compute the difference between the η^2 of the two Granger causal models. This is the effect size of the Granger causality. This is represented on the x-axis of the figure 4.
   (e) Compute the difference between the Granger causal model with higher η^2 and the η^2 of the correlational model. This is the significance of the Granger causality. This is represented on the y-axis of the figure 4.
(2) Once we have the effect size and the significance of the Granger causality for each group, plot them on a Cartesian-coordinate system (e.g., Figure 4).
(3) Next, we can compare the effect sizes of the Granger causality across the two levels of gains.

For additional details about Granger causality with more than one participant/group, please see previously published works in

[1]One could also use the greyscale image for computing the entropy of each frame in the screen recording. In this paper, we have used all the three channels for having a more accurate value of the entropy than while using a single greyscale image.

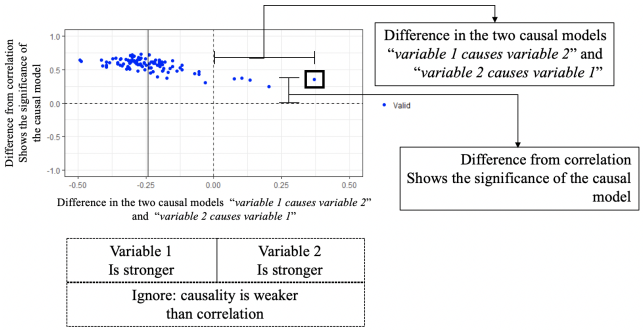

Figure 4: A visualization to summarize causality results for multiple participants. For each participant we calculate the difference between the two causal models ('x causes y' and 'y causes x') and the respective difference between causal and correlational models.

the context of learning technologies and HCI [94, 112]. The aforementioned steps were followed to identify the causal relations of children's JVA and information flow (RQ1) and JES and information flow (RQ2). Finally, to identify potential differences on the causal relations between children groups with low and high performance (RQ3), we employed a Wilcoxon test. Wilcoxon test allows us to compare the strength of the causal relations between groups. We used a non-parametric test because there is neither theoretical nor practical basis for assuming that effect sizes would follow a known statistical distributions (e.g., Gaussian, Poisson, Student-t).

4 RESULTS

We checked the bias of the number of members in a team on all the measurements (dyads vs. triads). We did not find any significant difference between the dyads and triads. In the next subsections we will present the results base don the causal analyses between: 1) JVA and information flow of the screen (RQ1), and JES and information flow of the screen (RQ2). Further, we analyse these causal relation with respect to the children's performance (RQ3). Before presenting the results, it is important to explain how the figures can be read in this section. Each figure in this section shows the strength (effect size) of the causal relation on the x-axis and the significance of the causal relation on the y-axis. Each point on the figures represents one group. The colour of the point depicts the high/low performance levels. Any point with negative y-axis should be ignored. Further, the positive y-axis on each figure is divided into two vertical halves. Any point in the left half shows that the information flow is causing the respective measurement (JVA or JES); and any point on the right half shows that the respective measurement is causing the information flow.

4.1 Information flow and JVA

We investigated whether the information flow is controlled by the children's JVA or the other way round. For this, we analysed the relation based on Granger causality between information flow and JVA. When considering both high and low performing groups, the results depict that JVA causes the information flow (the vertical lines in all the panels of figure 5 show the mean effect size for the whole sample). From all the panels of Figure 5 we can see the mean of the effect size, i.e., the difference between the two causal models: 1) the information flow Granger-causes the JVA, 2) the JVA Granger-causes the information flow. The mean effect sizes for the three phases are 0.19 (SD = 0.15), 0.25 (SD = 0.20), 0.30 (SD = 0.25). This indicates that children's JVA Granger-causes the information flow (RQ1). **Therefore, in the context of collaborative coding activities the way children collaborate (i.e., JVA) drives how the information is presented on the screen (i.e., information flow).**

When we compare the effect size of the Granger causality between the teams with high and low gains (RQ3), we observe that for all the three gains the effect size of JVA Granger causing the information flow for high gain teams was significantly higher than that for the low gain teams. **Meaning that the groups of children that performed high, also maintained a JVA that was a strong driver of the information flow. On the other hand this causality is weak for the low performing group.** Moreover, this difference is consistently increasing as children progress during the activity (i.e., the gain children obtained during the first 45 minutes (first gain) (W = 120, p < .00001) to the second 45 minutes (second gain) (W = 122, p < .00001) to the third 45 minutes (third gain) (W = 124, p < .00001)).

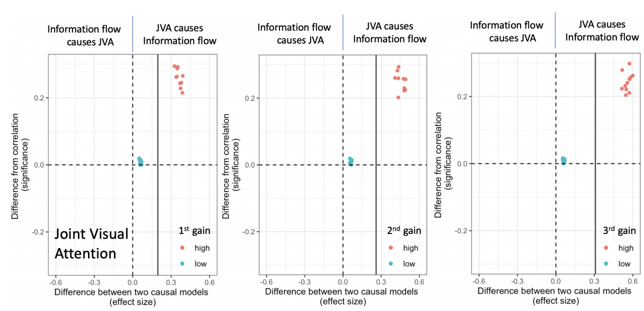

Figure 5: Results from analyzing the relation between Information flow and JVA.

4.2 Information flow and JES (RQ2 and RQ3)

4.2.1 Boredom. We investigated whether information flow is controlled by children's boredom JES or the other way round. For this, we analysed the relation based on Granger causality between information flow and boredom JES. When considering both high and low performing groups, the results depict that boredom JES causes the information flow (the vertical lines in all the panels of figure 6 show the mean effect size for the whole sample). From all the panels of Figure 6 we can see the mean of the effect size, i.e., the difference between the two causal models: 1) the information flow Granger-causes the boredom JES, 2) the boredom JES Granger-causes the information flow. The mean effect sizes for the three phases are -0.26 (SD = 0.31), -0.25 (SD = 0.17), -0.14 (SD = 0.18). This indicates that in formation flow Granger-causes children's boredom JES (RQ2). **Therefore, in the context of collaborative coding activities how the information is presented on the screen (i.e., information flow) drives joint emotional state of boredom.**

When we compare the effect size of the Granger causality between the teams with high and low gains (RQ3), we observe that for all the three gains the effect size of information flow Granger causing the boredom JES for low gain teams was significantly higher than that for the high gain teams.**Meaning that the groups of children that performed low, also had information flow as a strong driver of their boredom JES.** Moreover, this difference is consistently increasing as children progress during the activity (i.e., the gain children obtained during the first 45 minutes (first gain) (W = 117, p < .00001) to the second 45 minutes (second gain) (W = 121, p < .00001) to the third 45 minutes (third gain) (W = 122, p < .00001)).

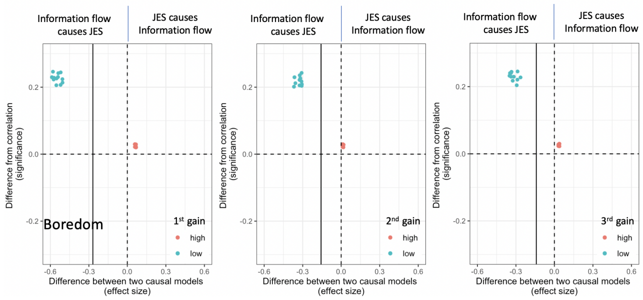

Figure 6: Results from analyzing the relation between Information flow and boredom JES

4.2.2 *Confusion.* We investigated whether information flow is controlled by children's confusion JES or the other way round. For this, we analysed the relation based on Granger causality between information flow and confusion JES. When considering both high and low performing groups, the results show confusion JES causing the information flow (the vertical lines in all the panels of figure 7 show the mean effect size for the whole sample). From all the panels of Figure 7 we can see the mean of the effect size, i.e., the difference between the two causal models: 1) the information flow Granger-causes confusion JES, 2) confusion JES Granger-causes the information flow. The mean effect sizes for the three phases are 0.33 (SD = 0.25), 0.18 (SD = 0.17), 0.16 (SD = 0.13). This indicates that children's JVA Granger-causes the information flow (RQ1). **Therefore, in collaborative coding context, the way children feel confused together (confusion JES) drives how information is presented on the screen (i.e., information flow).**

When we compare the effect size of the Granger causality between the teams with high and low gains (RQ3), we observe that for all the three gains the effect size of confusion JES Granger causing the information flow for low gain teams was significantly higher than that for the high gain teams. **Meaning that the groups of children that performed low, experienced confusion together, and that was a strong driver of the information flow.** Moreover, this difference is consistently increasing as children progress during the activity (i.e., the gain children obtained during the first 45 minutes (first gain) (W = 114, p < .00001) to the second 45 minutes (second gain) (W = 110, p < .00001) and the third 45 minutes (third gain) (W = 110, p < .00001)).

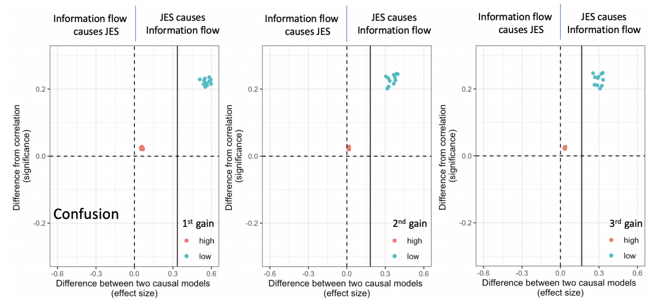

Figure 7: Results from analyzing the relation between Information flow and confusion JES.

4.2.3 *Frustration.* We investigated whether the information flow is controlled by children's frustration JES or the other way round. For this, we analysed the relation based on Granger causality between information flow and frustration JES. When considering both high and low performing groups, the results show frustration JES causing information flow (the vertical lines in all the panels of figure 8 show the mean effect size for the whole sample). From all the panels of Figure 8 we can see the mean of the effect size, i.e., difference between the two causal models: 1) the information flow Granger-causes frustration JES, 2) frustration JES Granger-causes the information flow. The mean effect sizes for the three phases are -0.31 (SD = 0.32), -0.26 (SD = 0.24), -0.10 (SD = 0.13). This indicates that information flow Granger-causes children's frustration JES (RQ2). **Therefore, in the context of collaborative coding activities how the information is presented on the screen (i.e., information flow) drives joint emotional state of frustration.**

When we compare the effect size of Granger causality between teams with high and low gains (RQ3), we observe that for all the three gains the effect size of information flow Granger causing the frustration JES for low gain teams was significantly higher than that for the high gain teams.**Meaning that the groups of children that performed low, also had information flow as a strong driver of their frustration JES.** Moreover, this difference is consistently increasing as children progress during the activity (i.e., the gain children obtained during the first 45 minutes (first gain) (W = 120, p < .00001) to the second and third 45 minutes (second and third gains) (W = 121, p < .00001).

5 DISCUSSION

5.1 Interpretations of Results

For this discussion, we will refer to "Granger cause/causal/causality" as "cause/causal/causality". Our results are consistent for all three phases of coding activity, where we calculated the gain between consecutive code evaluation based on the computational thinking components. The results show that there are clear differences based on the coding performance gain. We observe that for high gain groups JVA causes the information flow and there is a weak causal relation between the JES (boredom, confusion, and frustration) and information flow. On the other hand, for the low gain groups, we observe a weak causal relation between JVA and information flow. Moreover, for the low gain teams the information flow causes boredom JES and frustration JES; while the confusion JES causes the information flow.

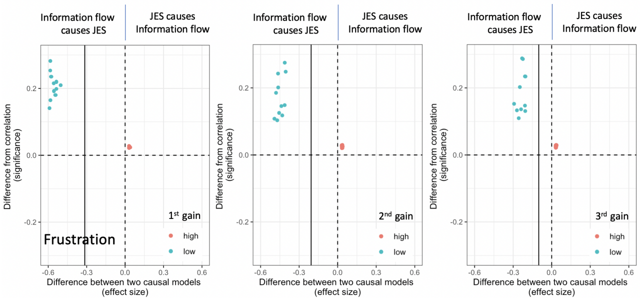

Figure 8: Results from analyzing the relation between Information flow and Frustration JES

One of the reasons that JVA causes the information flow for the high gain teams might be the fact that these teams produce more code when they have a high JVA level compared to when their JVA levels are low. JVA is found to be correlated with mutual understanding [84] and high collaboration quality [49]. When teams have high JVA they can be said to "be together" [91] and this could be one of the reasons that their code is of high quality, which is not the case in the low performing teams (where children have low JVA). Another important observation from the causal relation between JVA and information flow is that the difference between the strength of "JVA causing information flow" for high and low performance teams is increasing from the first phase to the second and the third phases. The primary reason for this could be based on theory of "convergent conceptual change" [80, 93, 108]. When collaborating peers are performing well in a given learning task, their conceptual models undergo a change such that the difference in their mental models about the problem is reduced than that of the beginning of the task [80]. This might be the case with the high performing teams; we observe that as the collaboration proceeds, such teams produce more content while being in a highly convergent (high JVA) state and less content while they are in a divergent (low JVA) state. The teams that do not show evolution of this behaviour end up performing poorer than the teams which achieve a converging conceptual change. This result is in line with related research [93, 108] which is primarily based on correlational methods.

Concerning JES, we observe that the low performing teams show strong causal relations with the information flow of the screen and not the high performing students. The emotions used in this paper were frustration, boredom and confusion, because they were the most prominent ones across different studies [5, 26, 27]. For low performing teams, the confusion JES causes the information flow of the screen, while the information flow causes the frustration JES and boredom JES. One reason for these results could be that when the peers in a low performing team are confused together, they try and produce a significant amount of content. It has been shown that teams experience confusion when their mental model about the problem/conversation at hand conflicts with the current state of the problem/conversation [16, 29]. Therefore, when teams are confused, it is unlikely that they produce high quality code and get a low performance rating. Once the code produced by the teams is of low quality, at a later stage of collaboration, the information content of the screen might cause boredom and/or frustration. This might induce the "vicious cycle" [29] of confusion phases of producing low quality code and that in turn causes boredom/frustration. It has been shown that in the case of boredom, students tend to "game the system" [4, 20], and in case of frustration their behaviour is "outside the locus of their control" [61]. In either of the cases the level of engagement with the problem at hand is deemed to be low [19]. Such situations should be avoided and the students should be supported to prevent these "vicious cycles".

In a nutshell, our results show that the teams that produce code when they are in a convergent state (high JVA) are the high performing teams. The low performing teams are the ones that produce code when the peers are confused and later bored/frustrated due to the produced code (poor quality) in earlier phases. These results, based on causal inference, provide us with a unique opportunity to design feedback/scaffolding tools that can aid the students to avoid the frustrating and boring episode; and at the same time, encourage them to work together. Moreover, we extend the previous findings about the socio-cognitive and affective processes, and how they relate to collaborative performance.

5.2 Implications for Research

The causal relation between JVA and information flow for the two performance levels (JVA causes information flow for high performing teams and no causal relation between JVA and information flow for low performing teams), extends the current knowledge about the relation between JVA and collaborative performance. In several studies, JVA was shown to be indicative of the "mutual understanding" between collaborators [49, 67, 86]. JVA was found to be correlated with collaborative learning performance [83, 92], including collaboration among children [73]. Further, JVA was also found to be correlated with the artefact quality [91], reason being the higher "mutual understanding" among the peers [66, 96]. Our results show that high performing teams add more content (increased information flow) in the moments of high-quality collaboration (increased JVA), which can not be said for the low performing students. Therefore, we extend the correlational equivalence of mutual understanding and JVA by showing that it is not the high performance that is "related" to high JVA, but it is the information flow that is "caused" by JVA for high performing teams. Therefore, we argue that the future research should consider the interplay performance and information presented on the stimulus to analyse JVA. Furthermore, since there is more information embedded in a causal relation than that in a correlation-based inference, there should also be a push for more causal inferences based studies.

When it comes to the facial expressions and/or the emotions in regards to collaborative learning and coding scenarios, our results extend the current state of knowledge. Recent research has shown that there is a clear relation between performance and emotions such as, frustration, confusion and boredom [5, 27, 47, 90, 118]. In certain studies evaluating the relation between the emotions and learning performance show a positive correlation between confusion and collaborative learning outcomes [29, 30]; while others report a negative correlation between collaborative learning outcomes and emotions such as, frustration and boredom [5, 28, 87]. Our results show that for low performing teams information flow causes boredom and frustration, while confusion causes information flow. Whereas, for high performing teams only a weak causation could be found. Our results extend the correlational knowledge-base with a causality-based knowledge; providing certain methodological implications for research in this area, and driving future works that will go beyond correlations and account for the causal relationships. Moreover, during collaboration among children, there has been some evidence that the JES is correlated to collaborative outcomes and experiences [98]. Our results not only extend such results by showing the causality between JES and information flow, but also explore the JES with respect to individual emotions (frustration, confusion, boredom). Exploring JES for individual emotions/expressions is important because if doing it without considering the actual emotion/expression, we are only considering children's similarity of emotions but not which emotions are they exhibiting [98]. This additional knowledge when combined with the causal information extends the framework of affective equilibrium for complex learning [29] (involving confusion, frustration, and boredom). This is achieved by adding the additional notion of causality with respect to the information flow as an external that might trigger impasses and affective disequilibrium [102].

Furthermore, When it comes to the choice of the method to analyze the causality between the different pairs of measurements. In this paper, we used the definition of causality provided by Granger. There are three other methods that could be used to show the causality between different variables: 1) Structured Equation Modelling (SEM, [31]) 2) Cross-convergent mapping (CCM, [104] intervention experiment where the hypothesised "cause" is controlled and the hypothesised "effect" is measured [88]. SEM does not necessarily contain the information required to consider a causal relationship. Statistically speaking, testing a SEM is not a test for causality. However, there are certain mathematical formulation under which SEM can be used for causal inference [103]. Bollen and Pearl [9] provide a detailed account describing how SEM should not be used for modelling causal relations between variables. The second method, CCM, is useful in cases where the time series is stationary (i.e., mean and variance of the variable do not change over time) and non-linear (i.e., there is no auto-correlation in the time series). Eye-tracking data is stationary (which can be tested using the Ljung-Box test) but auto-correlated (where users look at current time instance vastly depends on where they were looking at previous instances). This is why CCM is not an adequate method for such data. CCM was used in recent contributions [82, 112]. Finally, in the case of identifying causal relations between two variables through an experimental or pseudo-experimental setup, such setups are typically costly or require an extensive duration in order to identify the cause-effect relationship between the two variables in question [12]. Moreover, it has also been shown that for longer time series data the Granger causality outperforms other contemporary methods [120].

5.3 Implications for Design

Better ACLS can be designed by using the causal relations than correlational results because causality accounts for the information exchange and the underlying processes that are responsible for the two variables. From our results, there are two clear ways of providing adaptive support to the teams based on the two causal relations. Both the support types depend on the fact that we should look for the causal relation in real-time and find the moments when the causality changes the direction (or is no longer observed). On one hand, "JVA causing the information flow" should be supported and such behaviour should be encouraged; on the other hand "frustration and boredom JES causing the information flow" and "information flow causing confusion JES" should be prevented and appropriate scaffolding to avoid such behaviour should be provided to the children. We discuss these two cases in detail.

First, the strength and nature of the causal relationship between JVA and information flow differ across the performance levels has direct implications on the design of adaptive collaborative learning support (ACLS); e.g., a decrement in the strength of "JVA causing information flow" indicates potentially low performance. In this moment, the support system can trigger help which can increase the JVA among peers by directing their attention to a common (specially incomplete) part of the code. This is a way to ensure that the teams will add new content (increased information flow) with high JVA. Another use-case for providing JVA-based support, is during imbalanced division of labour. For example, if we observe that the level of JVA is decreasing (which will decrease the information flow), because there is only one child working on the code, we can direct the attention of non-contributing (or less contributing) child(ren) to the place where the contributing child(ren) is looking at. Gaze-aware feedback tools have been shown to have a positive impact on the collaborative learning performance [84], specially during coding scenarios [21, 22].

Second, we have shown that for low performing teams confusion JES causes the information flow, and information flow causes boredom JES and frustration JES. It is interesting that information flow causes the two JES (boredom and frustration) which are important to monitor for keeping an appropriate level of engagement in adaptive conditions [19]. Boredom is often mitigated by creating more challenging events during the interaction whereas, frustration is negotiated by providing content-based hints to make the content easier to understand. Based on the current state-of-the-art in affective means to support educational activities, both frustration and confusion might be beneficial for learning (confusion, [30]; frustration, [5]), while boredom can be detrimental for learning outcomes [28]. In this paper, we are considering the JES of theses emotions and suggest that we should avoid the situations where all members in a team have similar levels of these emotions. Baker et al [20] argue that boredom and confusion should be mitigated as soon as possible, while Mentis [63] argues that frustration might not require intervention at all. Pour et al. [76] suggest that boredom and confusion can be mitigated by directing children about "what needs to be fixed" and helping them understand the concepts, respectively. We propose that by providing similar support, the JES of boredom and confusion can be brought to a lower level so that the children do not produce content while experiencing these emotions. Further, from a study-design point of view, using webcam-based videos to estimate the gaze of the children addresses the ecological-validity threats that are imposed by the eye-tracking equipment. This also provides a method to have a low-fidelity but highly cost effective solution for the expensive eye-trackers.

Causal relations provide the instructional designers an opportunity to take appropriate actions so that the students access the information presented to them in an efficient and effective manner. By knowing the causal relations between the information flow and JVA/JES can provide instructional designers with guidelines/recommendations [82, 94, 104] to control the cause and modify the effect in the desired manner. The understanding of causal relationships could also help the teachers and researchers in avoiding unforeseen situations in the collaborative learning settings.

5.4 Limitations & Future Directions

The generalizability of our findings might be restricted by the specifically designed tasks; other tasks (e.g., video-based or game-based learning), or different representations of the same tasks (e.g., text-based coding) might affect the results. In particular, we used a specific task (i.e., collaborative coding with Scratch) to portray an active learning performance (i.e., the dependent variables). This tasks present a good case study in terms of the research questions, however, there are significant differences with other tasks that might affect the outcome. Therefore, we suggest the use of different tasks to portray learning performance in the future. Furthermore, the participants were from schools who showed interest in participating in our workshops, so other sampling methods could have been applied to attract more children leading to a more representative sample size of the target community. Furthermore, we only consider the whole interaction as one "session" to compute the causality between the information flow and the two measurements (JVA and JES). However, collaboration is dynamic and the causalities might change and provide more information if we consider smaller temporal windows. In the future, we will also aim to explore the changing nature of causal relations among multiple measurements to enhance our understanding. In addition, in this study we did not take into consideration the children's individual prior knowledge which might had affected the developed artifacts.

6 CONCLUSION

We presented causal relations between the information flow of the screen and the collaborative behaviour of ten dyads and ten triads from a coding workshop. On one hand, the high performing teams produce code when they have high levels of common ground and mutual understanding (JVA causes information flow) while on the other hand, the low performing teams produce code when they are confused (confusion JES causes information flow) and that in turn bores/frustrates them (information flow causes boredom JES and frustration JES). This might be the key explanation for teams' performance levels. We argue that there should be consistent efforts for exploring causality rather than conventional correlational analysis of behaviour and outcome because causal relations provide more information about the interplay of behavioural measurements than correlation.

7 SELECTION AND PARTICIPATION OF CHILDREN

All the participants of the study were students from the Norwegian University of Science and Technology (Trondheim, Norway) region whose teachers have applied to participate in our workshops as an out-of school activity. Studies took place at the university campus in specially designed rooms. Data related to the study were collected after permission from the national Data Protection Official for Research, following all the regulations and recommendations for research with children. A researcher contacted the teacher and the legal guardian of each child to get a written consent that gave permission for the data collection. The children were informed about the data collection process and their participation in the study was completely voluntary. They could withdraw their consent for the data collection at any time without affecting their participation in the coding activity.

REFERENCES

[1] Paul D Allopenna, James S Magnuson, and Michael K Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language* 38, 4 (1998), 419–439.

[2] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. 2016. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science* 6 (2016).

[3] Nicola Ancona, Daniele Marinazzo, and Sebastiano Stramaglia. 2004. Radial basis function approach to nonlinear Granger causality of time series. *Physical Review E* 70, 5 (2004), 056221.

[4] Ryan Shaun Baker, Albert T Corbett, Kenneth R Koedinger, and Angela Z Wagner. 2004. Off-task behavior in the cognitive tutor classroom: when students" game the system". In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 383–390.

[5] Ryan SJd Baker, Sidney K D'Mello, Ma Mercedes T Rodrigo, and Arthur C Graesser. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies* 68, 4 (2010), 223–241.

[6] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. 2016. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1–10. https://doi.org/10.1109/WACV.2016.7477553.

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

[8] Herbert Bless. 2000. The interplay of affect and cognition: The mediating role of general knowledge structures. (2000).

[9] Kenneth A Bollen and Judea Pearl. 2013. Eight myths about causality and structural equation models. In *Handbook of causal analysis for social research*. Springer, 301–328.

[10] Neus Capdeferro and Margarida Romero. 2012. Are online learners frustrated with collaborative learning experiences? *The International Review of Research in Open and Distributed Learning* 13, 2 (2012), 26–44.

[11] Areti Chalkiadaki. 2018. A systematic literature review of 21st century skills and competencies in primary education. *International Journal of Instruction* 11, 3 (2018), 1–16.

[12] Daniel F Chambliss and Russell K Schutt. 2018. *Making sense of the social world: Methods of investigation*. Sage Publications.

[13] Carol Chan, Jud Burtis, and Carl Bereiter. 1997. Knowledge building as a mediator of conflict in conceptual change. *Cognition and instruction* 15, 1 (1997), 1–40.

[14] Sue-Jen Chen and Edward J Caropreso. 2004. Influence of personality on online discussion. *Journal of Interactive Online Learning* 3, 2 (2004), 1–17.

[15] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding. 2004. Analyzing multiple nonlinear time series with extended Granger causality. *Physics Letters A* 324, 1 (2004), 26–35.

[16] Geraldine Clarebout and Jan Elen. 2001. The ParlEuNet-project: Problems with the validation of socio-constructivist design principles in ecological settings. *Computers in Human Behavior* 17, 5-6 (2001), 453–464.

[17] Herbert H Clark and Keith Brown. 2006. Context and common ground. *Concise Encyclopedia of Philosophy of Language and Linguistics* (2006), 85–87.

[18] Herbert H Clark, Robert Schreuder, and Samuel Buttrick. 1983. Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior* 22, 2 (1983), 245–258.

[19] Mihaly Csikszentmihalyi. 1997. Flow and the psychology of discovery and invention. *HarperPerennial, New York* 39 (1997).

[20] Ryan SJ d Baker, Albert T Corbett, Ido Roll, and Kenneth R Koedinger. 2008. Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction* 18, 3 (2008), 287–314.

[21] 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.

[22] Sarah D'Angelo and Darren Gergle. 2018. An Eye For Design: Gaze Visualizations for Remote Collaborative Work. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 349.

[23] Agnes Delaborde, Marie Tahon, Claude Barras, and Laurence Devillers. 2009. A Wizard-of-Oz game for collecting emotional audio data in a children-robot interaction. In *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*. ACM, 5.

[24] Mingzhou Ding, Steven L Bressler, Weiming Yang, and Hualou Liang. 2000. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biological cybernetics* 83, 1 (2000), 35–45.

[25] Mingzhou Ding, Yonghong Chen, and Steven L Bressler. 2006. 17 Granger causality: basic theory and application to neuroscience. *Handbook of time series analysis: recent theoretical developments and applications* 437 (2006).

[26] Sidney D'Mello. 2013. A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology* 105, 4 (2013), 1082.

[27] Sidney D'Mello and Art Graesser. 2010. Modeling cognitive-affective dynamics with Hidden Markov Models. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 32.

[28] Jason E Dowd, Ives Araujo, and Eric Mazur. 2015. Making sense of confusion: Relating performance, confidence, and self-efficacy to expressions of confusion in an introductory physics class. *Physical Review Special Topics-Physics Education Research* 11, 1 (2015), 010107.

[29] Sidney D'Mello and Art Graesser. 2012. Dynamics of affective states during complex learning. *Learning and Instruction* 22, 2 (2012), 145–157.

[30] Sidney D'Mello, Blair Lehman, Reinhard Pekrun, and Art Graesser. 2014. Confusion can be beneficial for learning. *Learning and Instruction* 29 (2014), 153–170.

[31] Jeffrey R Edwards and Richard P Bagozzi. 2000. On the nature and direction of relationships between constructs and measures. *Psychological methods* 5, 2 (2000), 155.

[32] Barbara L Fredrickson. 1998. What good are positive emotions? *Review of general psychology* 2, 3 (1998), 300.

[33] Fernando Garcia-Sanjuan, Sandra Jurdi, Javier Jaen, and Vicente Nacher. 2018. Evaluating a tactile and a tangible multi-tablet gamified quiz system for collaborative learning in primary education. *Computers & Education* 123 (2018), 65–84.

[34] Darren Gergle and Alan T Clark. 2011. See what i'm saying?: using Dyadic Mobile Eye tracking to study collaborative reference. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 435–444.

[35] Soudeh Ghorbani and Brighten Godfrey. 2014. Towards correct network virtualization. *ACM SIGCOMM Computer Communication Review* 44, 4 (2014).

[36] Michail N Giannakos, Konstantinos Chorianopoulos, Kori Inkpen, Honglu Du, and Paul Johns. 2013. Understanding children's behavior in an asynchronous video-mediated communication environment. *Personal and ubiquitous computing* 17, 8 (2013), 1621–1629.

[37] Michail N Giannakos, Michael S Horn, Janet C Read, and Panos Markopoulos. 2020. Movement forward: The continued growth of Child-Computer Interaction research.

[38] Rainer Goebel, Alard Roebroeck, Dae-Shik Kim, and Elia Formisano. 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic resonance imaging* 21, 10 (2003), 1251–1261.

[39] Rafael C Gonzalez and Richard E Woods. 2007. Image processing. *Digital image processing* 2, 1 (2007).

[40] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* (1969), 424–438.

[41] Zenzi M Griffin and Jordan C Davison. 2011. A technical introduction to using speakers' eye movements to study language. *The mental lexicon* 6, 1 (2011), 53–82.

[42] P EKMAN-WV FRIESEN-JC HAGER. 2002. Facial Action Coding System. The Manual On CD ROM.

[43] Jeongyun Han, Kwan Hoon Kim, Wonjong Rhee, and Young Hoan Cho. 2021. Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Computers & Education* 163 (2021), 104041.

[44] Natalie Harrold, Chek Tien Tan, Daniel Rosser, and Tuck Wah Leong. 2014. CopyMe: an emotional development game for children. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 503–506.

[45] Wolfram Hesse, Eva Möller, Matthias Arnold, and Bärbel Schack. 2003. The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *Journal of neuroscience methods* 124, 1 (2003), 27–44.

[46] Craig Hiemstra and Jonathan D Jones. 1994. Testing for linear and nonlinear Granger causality in the stock price-volume relation. *The Journal of Finance* 49, 5 (1994), 1639–1664.

[47] Jennifer A Hunter and John D Eastwood. 2021. Understanding the relation between boredom and academic performance in postsecondary students. *Journal of Educational Psychology* 113, 3 (2021), 499.

[48] Hanna Järvenoja, Jonna Malmberg, Tiina Törmänen, Kristiina Mänty, Eetu Haataja, Sara Ahola, and Sanna Järvelä. 2020. A Collaborative Learning Design for Promoting and Analyzing Adaptive Motivation and Emotion Regulation in the Science Classroom. In *Frontiers in Education*, Vol. 5. Frontiers, 111.

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

[50] Patrick Jermann, Marc-Antoine Nüssli, and Weifeng Li. 2010. Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*. British Computer Society, 36–44.

[51] Samantha Jiménez, Reyes Juárez-Ramírez, Víctor H Castillo, Guillermo Licea, Alan Ramírez-Noriega, and Sergio Inzunza. 2018. A feedback system to provide affective support to students. *Computer Applications in Engineering Education* 26, 3 (2018), 473–483.

[52] Wayne Joerding. 1986. Economic growth and defense spending: Granger causality. *Journal of Development Economics* 21, 1 (1986), 35–40.

[53] Ashish Kapoor and Rosalind W Picard. 2005. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 677–682.

[54] Ryan A Kirk, Dave A Kirk, and Peter Pesheck. 2016. Decision making for complex ecosystems: a technique for establishing causality in dynamic systems. In *International Conference on Human-Computer Interaction*. Springer, 110–115.

[55] Kyungbin Kwon, Ying-Hsiu Liu, and LaShaune P Johnson. 2014. Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers & Education* 78 (2014), 185–200.

[56] Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, Ana Paiva, and Peter W McOwan. 2009. Designing a game companion for long-term social interaction. In *Proceedings of the International Workshop on Affective-aware Virtual Agents and Social Robots*. ACM, 10.

[57] Iolanda Leite, Rui Henriques, Carlos Martinho, and Ana Paiva. 2013. Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 41–48.

[58] X San Liang. 2018. Causation and information flow with respect to relative entropy. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, 7 (2018), 075311.

[59] Huimin Liu. 2008. Generalized additive model. *Department of Mathematics and Statistics University of Minnesota Duluth: Duluth, MN, USA* (2008).

[60] Chad N Loes and Ernest T Pascarella. 2017. Collaborative learning and critical thinking: Testing the link. *The Journal of Higher Education* 88, 5 (2017), 726–753.

[61] Lazarus Ndiku Makewa, Dorcas Gitonga, Baraka Ngussa, Samwel Njoroge, and Joshua Kuboja. 2014. Frustration factor in group collaborative learning experiences. *American Journal of Educational Research* 2, 11A (2014), 16–22.

[62] Bethany McDaniel, Sidney D'Mello, Brandon King, Patrick Chipman, Kristy Tapp, and Art Graesser. 2007. Facial features for affective state detection in learning environments. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 29.

[63] Helena M Mentis et al. 2007. Memory of frustrating experiences. *Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory, Eds. Diane Nahl and Dania Bilal* (2007), 197–210.

[64] Jesús Moreno-León, Gregorio Robles, and Marcos Román-González. 2015. Dr. Scratch: Automatic analysis of scratch projects to assess and foster computational thinking. *RED. Revista de Educación a Distancia* 46 (2015), 1–23.

[65] Jesús Moreno-León, Marcos Román-González, Casper Harteveld, and Gregorio Robles. 2017. On the automatic assessment of computational thinking skills: A comparison with human experts. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2788–2795.

[66] 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.

[67] Jennifer K Olsen, Daniel M Belenky, Vincent Aleven, and Nikol Rummel. 2014. Using an intelligent tutoring system to support collaborative as well as individual learning. In *International Conference on Intelligent Tutoring Systems*. Springer, 134–143.

[68] Jennifer K Olsen, Nikol Rummel, and Vincent Aleven. 2019. It is not either or: An initial investigation into combining collaborative and individual learning using an ITS. *International Journal of Computer-Supported Collaborative Learning* 14, 3 (2019), 353–381.

[69] Harmen Oppewal. 2010. Causal research. *Wiley International Encyclopedia of Marketing* (2010).

[70] Theodore Panitz. 1999. The Case for Student Centered Instruction via Collaborative Learning Paradigms. (1999).

[71] Sofia Papavlasopoulou, Michail N Giannakos, and Letizia Jaccheri. 2017. Empirical studies on the Maker Movement, a promising approach to learning: A literature review. *Entertainment Computing* 18 (2017), 57–78.

[72] Sofia Papavlasopoulou, Michail N Giannakos, and Letizia Jaccheri. 2019. Exploring children's learning experience in constructionism-based coding activities

through design-based research. *Computers in Human Behavior* 99 (2019), 415–427.

[73] Sofia Papavlasopoulou, Kshitij Sharma, Michail Giannakos, and Letizia Jaccheri. 2017. Using eye-tracking to unveil differences between kids and teens in coding activities. In *proceedings of the 2017 conference on interaction design and children*. 171–181.

[74] Reinhard Pekrun. 2006. The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review* 18, 4 (01 Dec 2006), 315–341. https://doi.org/10.1007/s10648-006-9029-9

[75] Daniel Perry and Cecilia Aragon. 2012. Measuring distributed affect in collaborative games. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*. ACM, 195–198.

[76] Payam Aghaei Pour, M Sazzad Hussain, Omar AlZoubi, Sidney D'Mello, and Rafael A Calvo. 2010. The impact of system feedback on learners' affective and physiological states. In *International Conference on Intelligent Tutoring Systems*. Springer, 264–273.

[77] Rachel Carlos Duque Reis, Seiji Isotani, Carla Lopes Rodriguez, Kamila Takayama Lyra, Patrícia Augustin Jaques, and Ig Ibert Bittencourt. 2018. Affective states in computer-supported collaborative learning: Studying the past to drive the future. *Computers & Education* 120 (2018), 29–50.

[78] Daniel C Richardson, Rick Dale, and Natasha Z Kirkham. 2007. The art of conversation is coordination. *Psychological science* 18, 5 (2007), 407–413.

[79] Elena Rondeau, Nicolas Larmonier, Thomas Pradeu, and Andreas Bikfalvi. 2019. Philosophy of Biology: Characterizing causality in cancer. *Elife* 8 (2019), e53755.

[80] Jeremy Roschelle and Stephanie D Teasley. 1995. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning*. Springer, 69–97.

[81] Sreecharan Sankaranarayanan, Siddharth Reddy Kandimalla, Sahil Hasan, Haokang An, Christopher Bogart, R Charles Murray, Michael Hilton, Majd Sakr, and Carolyn Rosé. 2020. Agent-in-the-loop: conversational agent support in service of reflection for learning during collaborative programming. In *International Conference on Artificial Intelligence in Education*. Springer, 273–278.

[82] Zhanna Sarsenbayeva, Gabriele Marini, Niels van Berkel, Chu Luo, Weiwei Jiang, Kangning Yang, Greg Wadley, Tilman Dingler, Vassilis Kostakos, and Jorge Goncalves. 2020. Does smartphone use drive our emotions or vice versa? A causal analysis. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15.

[83] Bertrand Schneider, Sami Abu-El-Haija, Jim Reesman, and Roy Pea. 2013. Toward collaboration sensing: applying network analysis techniques to collaborative eye-tracking data. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. ACM, 107–111.

[84] 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.

[85] Bertrand Schneider, Kshitij Sharma, Sebastien Cuendet, Guillaume Zufferey, Pierre Dillenbourg, and Roy Pea. 2016. Detecting collaborative dynamics using mobile eye-trackers. Singapore: International Society of the Learning Sciences.

[86] Bertrand Schneider, Megan Strait, Laurence Muller, Sarah Elfenbein, Orit Shaer, and Chia Shen. 2012. Phylo-Genie: engaging students in collaborative 'tree-thinking' through tabletop techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3071–3080.

[87] Karen E Seymour, Keri S Rosch, Alyssa Tiedemann, and Stewart H Mostofsky. 2020. The validity of a frustration paradigm to assess the effect of frustration on cognitive control in school-age children. *Behavior therapy* 51, 2 (2020), 268–282.

[88] William R Shadish, Thomas D Cook, Donald Thomas Campbell, et al. 2002. *Experimental and quasi-experimental designs for generalized causal inference/William R. Shedish, Thomas D. Cook, Donald T. Campbell*. Boston: Houghton Mifflin,.

[89] Suleman Shahid, Emiel Krahmer, Marc Swerts, and Omar Mubin. 2010. Child-robot interaction during collaborative game play: Effects of age and gender on emotion and experience. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. ACM, 332–335.

[90] Kaiqi Shao, Reinhard Pekrun, Herbert W Marsh, and Kristina Loderer. 2020. Control-value appraisals, achievement emotions, and foreign language performance: A latent interaction analysis. *Learning and Instruction* 69 (2020), 101356.

[91] Kshitij Sharma, Daniela Caballero, Himanshu Verma, Patrick Jermann, and Pierre Dillenbourg. 2015. Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. International Society of the Learning Sciences, Inc.[ISLS].

[92] Kshitij Sharma, Patrick Jermann, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2012. Gaze Evidence for different activities in program understanding. In *24th Annual conference of Psychology of Programming Interest Group*.

[93] Kshitij Sharma, Ioannis Leftheriotis, and Michail Giannakos. 2020. Utilizing Interactive Surfaces to Enhance Learning, Collaboration and Engagement: Insights from Learners' Gaze and Speech. *Sensors* 20, 7 (2020), 1964.

[94] Kshitij Sharma, Katerina Mangaroska, Niels van Berkel, Michail Giannakos, and Vassilis Kostakos. 2021. Information flow and cognition affect each other: Evidence from digital learning. *International Journal of Human-Computer Studies* 146 (2021), 102549.

[95] Kshitij Sharma, Jennifer K Olsen, Vincent Aleven, and Nikol Rummel. 2018. Exploring Causality Within Collaborative Problem Solving Using Eye-Tracking. In *European Conference on Technology Enhanced Learning*. Springer, 412–426.

[96] Kshitij Sharma, Jennifer K Olsen, Vincent Aleven, and Nikol Rummel. 2020. Measuring causality between collaborative and individual gaze metrics for collaborative problem-solving with intelligent tutoring systems. *Journal of Computer Assisted Learning* (2020).

[97] Kshitij Sharma, Sofia Papavlasopoulou, and Michail Giannakos. 2019. Coding games and robots to enhance computational thinking: How collaboration and engagement moderate children's attitudes? *International Journal of Child-Computer Interaction* 21 (2019), 65–76.

[98] Kshitij Sharma, Sofia Papavlasopoulou, and Michail Giannakos. 2019. Joint Emotional State of Children and Perceived Collaborative Experience in Coding Activities. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. ACM, 133–145.

[99] Geoffrey Smith. 2009. On the foundations of quantitative information flow. In *International Conference on Foundations of Software Science and Computational Structures*. Springer, 288–302.

[100] Samuel Spaulding, Goren Gordon, and Cynthia Breazeal. 2016. Affect-aware student models for robot tutors. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 864–872.

[101] Priyashri K Sridhar, Samantha WT Chan, and Suranga Nanayakkara. 2018. Going beyond performance scores: understanding cognitive-affective states in kindergarteners. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, 253–265.

[102] Nancy L Stein and Linda J Levine. 1990. Making sense out of emotion: The representation and use of goal-structured knowledge. *Psychological and biological approaches to emotion* (1990), 45–73.

[103] Rolf Steyer, Christof Nachtigall, Olivia Wüthrich-Martone, and Katrin Kraus. 2002. Causal regression models III: Covariates, conditional, and unconditional average causal effects. *Methods of Psychological Research Online* 7, 1 (2002), 41–68.

[104] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. 2012. Detecting causality in complex ecosystems. *science* 338, 6106 (2012), 496–500.

[105] Noorul Diana Sulaiman and Masitah Shahrill. 2015. Engaging collaborative learning to develop students' skills of the 21st century. *Mediterranean Journal of Social Sciences* 6, 4 (2015), 544.

[106] Kenji Suzuki. 2015. Social imaging technology to identify and represent social behaviors. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 907–908.

[107] Yuji Takano and Kenji Suzuki. 2014. Affective communication aid using wearable devices based on biosignals. In *Proceedings of the 2014 conference on Interaction design and children*. ACM, 213–216.

[108] Ping-Kee Tao and Richard F Gunstone. 1999. Conceptual change in science through collaborative learning at the computer. *International Journal of Science Education* 21, 1 (1999), 39–57.

[109] Giovanni Maria Troiano, Sam Snodgrass, Erinç Argımak, Gregorio Robles, Gillian Smith, Michael Cassidy, Eli Tucker-Raymond, Gillian Puttick, and Casper Harteveld. 2019. Is my game OK Dr. Scratch? Exploring programming and computational thinking development via metrics in student-designed serious games for STEM. In *Proceedings of the 18th ACM international conference on interaction design and children*. 208–219.

[110] Tzu-Wei Tsai, Hsiao Yu Lo, and Kai-Shao Chen. 2012. An affective computing approach to develop the game-based adaptive learning material for the elementary students. In *Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments*. ACM, 8–13.

[111] Jennifer Tsan, Collin F Lynch, and Kristy Elizabeth Boyer. 2018. "Alright, what do we need?": A study of young coders' collaborative dialogue. *International Journal of Child-Computer Interaction* 17 (2018), 61–71.

[112] Niels van Berkel, Simon Dennis, Michael Zyphur, Jinjing Li, Andrew Heathcote, and Vassilis Kostakos. 2020. Modeling interaction as a complex system. *Human–Computer Interaction* (2020), 1–27.

[113] Ester Van Laar, Alexander JAM Van Deursen, Jan AGM Van Dijk, and Jos De Haan. 2017. The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in human behavior* 72 (2017), 577–588.

[114] Kurt VanLehn, Stephanie Siler, Charles Murray, Takashi Yamauchi, and William B Baggett. 2003. Why do only some events cause learning during human tutoring? *Cognition and Instruction* 21, 3 (2003), 209–249.

[115] Erin Walker, Nikol Rummel, and Kenneth R Koedinger. 2011. Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning* 6,

2 (2011), 279–306.

[116] Genki Yamashita, Taro Kanno, and Kazuo Furuta. 2020. Interactive Method to Elicit Local Causal Knowledge for Creating a Huge Causal Network. In *International Conference on Human-Computer Interaction*. Springer, 437–446.

[117] Li Zhang. 2008. Metaphorical affect sensing in an intelligent conversational agent. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*. ACM, 100–106.

[118] Yingbin Zhang, Luc Paquette, Ryan S Baker, Jaclyn Ocumpaugh, Nigel Bosch, Gautam Biswas, and Anabil Munshi. [n.d.]. Can strategic behavior facilitate confusion resolution? The interplay between confusion and metacognitive strategies in Betty's Brain. ([n. d.]).

[119] S Sahand Mohammadi Ziabari and Jan Treur. 2018. An adaptive cognitive temporal-causal network model of a mindfulness therapy based on music. In *International Conference on Intelligent Human Computer Interaction*. Springer, 180–193.

[120] Cunlu Zou and Jianfeng Feng. 2009. Granger causality vs. dynamic Bayesian network inference: a comparative study. *BMC bioinformatics* 10, 1 (2009), 1–17.