Using Eye-Tracking Technology to Support Visual Coordination in Collaborative Problem-Solving Groups

Bertrand Schneider, Stanford University, schneibe@stanford.edu
Roy Pea, Stanford University, roypea@stanford.edu

Abstract: In this paper we present the results of an eye-tracking study on collaborative problem-solving dyads. Dyads remotely worked on contrasting cases to study how the human brain processes visual information. In one condition, dyads saw the gaze of their partner on the screen; in a control group, they did not have access to this information. Results indicate that this real-time mutual gaze perception intervention helped students achieve a higher quality of collaboration and a higher learning gain. Implications for supporting group collaboration are discussed.

Introduction

Joint attention is defined as "the tendency for social partners to focus on a common reference and to monitor one another's attention to an outside entity, such as an object, person, or event. [...] The fact that two individuals are simultaneously focused on the same aspect of the environment at the same time does not constitute joint attention. To qualify as joint attention, the social partners need to demonstrate awareness that they are attending to something in common" (Tomasello, 1995, pp. 86-87). Joint attention is fundamental to social coordination: young infants communicate emotions in a state of synchrony with their caregivers, in turn helping them achieve visual coordination when learning language (Stern, 1977). Parents use *deictic gestures* (i.e., pointing at a focus of interest to establish joint visual attention) to signal important features of the environment to their children (Bates et al., 1989). Professors and mentors teach by highlighting subtle nuances between students' and experts' conceptual understanding (Roth, 2001). Groups of students rely on coordination between their members to reach the problem solution (Barron, 2003), in turn influencing their level of abstract thinking (Schwartz, 1995).

We argue that the construction of perceptual joint attention rests significantly though not entirely (1) on two primary channels of communication: people can either point at things *physically* (i.e., using deictic gestures) or *verbally* (i.e., by describing the object of interest). These two mechanisms are not terribly efficient because misunderstanding can happen on several levels: verbally, communication is prone to misinterpretation from the receiver. This is likely to happen when experts are teaching novices, because novices are still learning the perceptual skills to isolate subtle features or patterns that separate them from experts. Physically, there is an extra step of taking the point of view of another person. From a spatial and social point of view, this is not a trivial mental operation (especially for children as demonstrated by Piaget in his studies of egocentrism and in more recent studies on the role of 'theory of mind' in human development; Leudar, Costall & Francis, 2004).

Previous work in CSCL used eye-trackers to study joint attention in collaborative learning situations. Richardson, Dale and Kirkham (2007) showed that common knowledge grounding positively influences the coordination of visual attention. Sangin (2009) studied pairs of students remotely working on a concept map and found evidence that knowledge awareness tools (i.e., displaying the level of expertise of each member of the dyad) was associated with a higher density of gaze-coupling and a higher quality of collaboration. Jermann, Nuessli, Mullins and Dillenbourg (2011) used synchronized eye-trackers to assess how programmers collaboratively work on a segment of code; they contrasted a good and bad dyad, and their results suggest that a productive collaboration is associated with high joint visual recurrence. Finally, Cherubini, Nuessli and Dillenbourg (2008) designed an algorithm that detects misunderstanding in a remote collaboration by using the distance between the gaze of the emitter and the receiver. Taken together, those evidences suggest that eye-trackers are a promising way to understand and predict the factors responsible for a high-quality collaboration.

Based on those studies, our goal is to develop new ways of supporting the establishment of *perceptual* joint attention (as opposed to cognitive, or social joint attention). We use eye-tracking technologies to share users' gaze during collaborative learning. More specifically, our first attempt involves dyads studying contrasting cases (Schwartz & Bransford, 1998). Our hypotheses are as follows: first, we expect dyads with access to their partner's gaze to have a higher quality of collaboration because such information will disambiguate their focus of attention and better enable "common ground" for learning conversations (Clark & Brennan, 1991). Secondly, we assume that a better collaboration will positively impact participants' learning gain (Barron, 2003).

General Description of the Experiment

The experiment had three distinct steps: during the first 12 minutes, dyads worked on 5 contrasting cases in neuroscience. They had to collaboratively explain how visual information is processed by the human brain based on what they have learned from the models described in Figure 1. They then read a text on the same topic for 12 minutes. Finally, they answered a learning test with questions on the terminology used, concepts taught and questions in which they needed to transfer their knowledge to a new situation.

Methods

Participants

Participants were 42 college-level students from a community college (average age 23.0, SD = 8.3; 28 females, 14 males). Dyads were randomly assigned to the two experimental conditions: the treatment group was in the "visible-gaze" condition (N = 24) with 16 females and 8 males; the control group was in the "no-gaze" condition, with 16 females and 8 males (N = 20). There was no significant difference in terms of GPA between the two conditions: $F(1,36) = 0.29$, $p = 0.59$ (for "visible-gaze": mean = 3.09, SD = 0.87; for "no-gaze": mean = 3.22, SD = 0.59). All participants were taking an introductory class in psychology and were required to participate in an experiment as part of their course.

Material

During the first step of the experiment, dyads worked on the contrasting cases shown in Fig. 1. To force collaboration, the answer of lesion 1 (top left) was visible only to the first member of the group while the answer of lesion 6 (top right) was shown only to the second member of the dyad. This kind of "jigsaw" method is commonly used to assure that one member of the dyad does not solve the problem alone (Aronson et al., 1978). The text used in the next step is available online(2). Finally the learning test contained 15 questions: 5 terminology questions (participants were asked to provide the name of a specific brain region or pathway), 5 conceptual questions (participants had to predict the effect of a specific lesion), and 5 transfer questions (subjects had to use their new knowledge to solve a vignette; e.g. "patient X is likely to have a lesion in region Y of the brain; should he be allowed to drive?"). All the material was exactly identical in the two conditions.

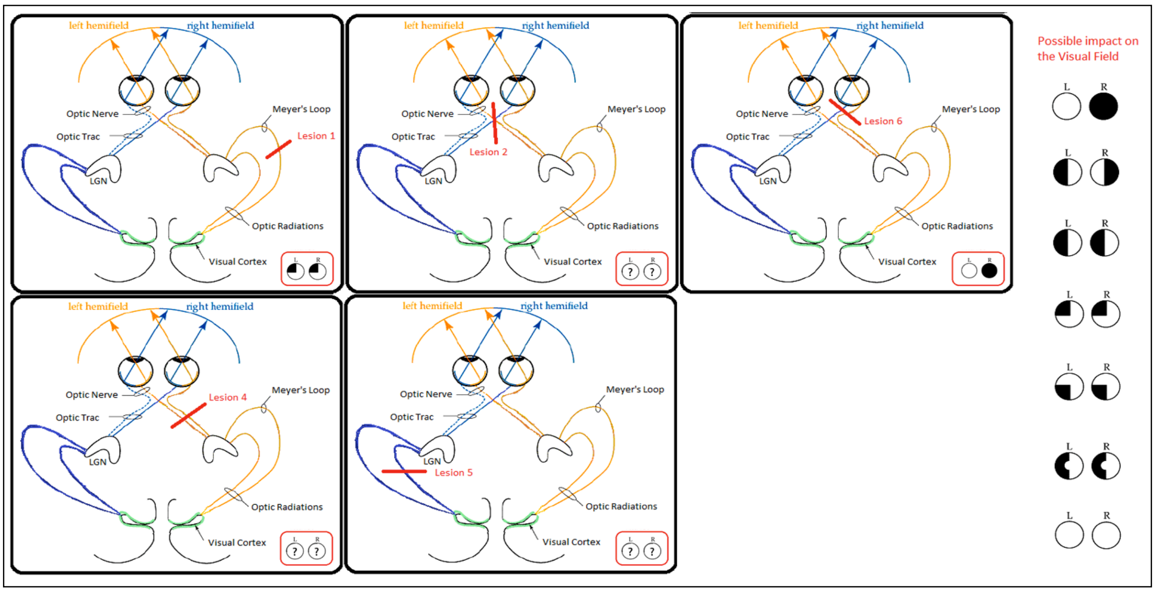

Figure 1. The dyads worked on the five contrasting cases above. Possible answers are shown on the right side. Answers of two cases (#1, top left and #6, top right) were given to subjects. Participants had to solve the three remaining cases (#2, top middle and #4, #5 bottom left and right of the screen).

Design

We used a between-subjects design with two conditions. In the "visible-gaze" condition, dyads were able to see the gaze of their partner on the screen. In the "no-gaze" condition, they could not. In the former condition, the gaze was only visible during the first step of the experiment (i.e., when dyads had to solve contrasting cases).

Procedure

Upon their arrival, participants were welcomed and thanked for their participation. The experimenter then explained that they would need to collaborate and suggested that they introduce themselves to their partner.

They were also told that each member of the dyad would be in a different room but would be able to communicate via a microphone. The experimenter explained that they would learn basic concepts in neuroscience, and he described the structure of the experiment (12 minutes of contrasting cases, 12 minutes of reading a text and as much time as needed for the learning test). Each participant then followed the experimenter to different rooms, where he calibrated their personal eye-tracker. The contrasting cases were then briefly presented to each participant and the experimenter ensured they understood the goal of the task. Subjects then worked on the contrasting cases and tried to determine how different lesions affected the brain's visual field. After 12 minutes, the screen automatically switched to a text explaining how the human brain processes visual information. The experimenter told participants they should read the text individually and then discuss it with their partner. After 12 minutes, the screen automatically switched to the learning test. The experimenter then told the subjects to individually complete the test and stopped the audio link. Participants took as much time as they needed for completion. They were then debriefed as the experimenter explained the goal of the study.

Measures

Because no participant had previous knowledge in neuroscience, learning gains were computed from the final learning test, which had three sub-dimensions: *conceptual questions* (predicting the effect of a particular lesion), *terminology* (naming brain regions or neural pathways) and *transfer questions* (solving a word problem using the concepts learned). The quality of collaboration was rated using dimensions developed in Meier, Spada and Rummel (2007), who assessed collaboration on a 5-point scale across 9 dimensions (sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, task management, technical coordination, reciprocal interaction and individual task orientation). The evaluation of this rating scheme demonstrated a high inter-rater reliability, consistency and validity, which renders it an appropriate tool for assessing collaboration. Finally, we categorized each participant as being a "follower" or a "leader" in the activity. We acknowledge subjects are likely to shift roles while solving contrasting cases. This measure can be considered as an aggregate estimation over the whole activity of the dyad's dynamic profile. We used indicators to categorize the dyad's members: 1) who starts the discussion when the experimenter leaves, 2) who speaks most, 3) who manages turn-taking (e.g., by asking "what do think?", "how do you understand this part of the diagram?"), and 4) who decides the next focus of attention (e.g., "so to summarize, our answers are [...]. I think we need to spend more time on diagram x"). We also collected eye-tracking data during the experiment: approximately 30 data-points per second were captured for each participant. This gave us ~1'000'000 gaze points in total. Within those measurements, we also collected participants' pupil size.

Results

In this section, we compare main effects for learning gains and collaboration scores across our two experimental groups. We then characterize the dyads of our experiments in terms of their gaze patterns by analyzing our eye-tracking data. We also compare process variables in terms of their predictive effect as mediators. Finally, we conclude by conducting a small qualitative analysis of two dyads (one from each experimental group) to suggest mechanisms explaining the main effects found.

Learning and Collaboration

As predicted, we found that participants in the "visible-gaze" group outperformed the dyads in the "no-gaze" condition for the total learning gain: $F(1,40) = 7.81$, $p < 0.01$. For the sub-dimensions, they also scored higher on the transfer questions $F(1,40) = 4.47$, $p < 0.05$. The difference would likely to be significant with a larger sample for the terminology questions $F(1,40) = 3.59$, $p = 0.065$ and for the conceptual questions $F(1,40) = 2.11$, $p = 0.154$, since the effect sizes are between medium and large (Cohen's d is 0.62 and 0.5, respectively).

The treatment group ("visible gaze") also had a higher quality of collaboration as measured by Meier, Spada and Rummel's (2007) rating scheme (the total score is an average across the 9 sub-dimensions described in the "measure" section): $F(1,19) = 11.73$, $p < 0.01$ (mean for the treatment group = 0.89, SD = 0.48; mean for the control group = -0.08, SD = 0.79). More specifically, those dyads were better at *sustaining mutual understanding* ($F(1,19) = 5.15$, $p < 0.05$), *pooling information* ($F(1,19) = 7.53$, $p < 0.05$), *reaching consensus* ($F(1,19) = 22.57$, $p < 0.001$) and *managing time* ($F(1,19) = 4.98$, $p < 0.05$). A second judge double-coded 20% of the video data; inter-reliability index using Krippendorff's alpha was 81.63%. An alpha higher than 80% is considered as a reliable agreement between judges (Hayes & Krippendorff, 2007).

Additionally we categorized each member of the dyad as "leader" and "follower" (Fig. 1). Interestingly we found an interaction effect between those two factors (experimental conditions and individuals' status) on the total learning score: $F(1,38) = 5.29$, $p < 0.05$. Followers learnt significantly more when they could see the gaze of the leader on the screen.

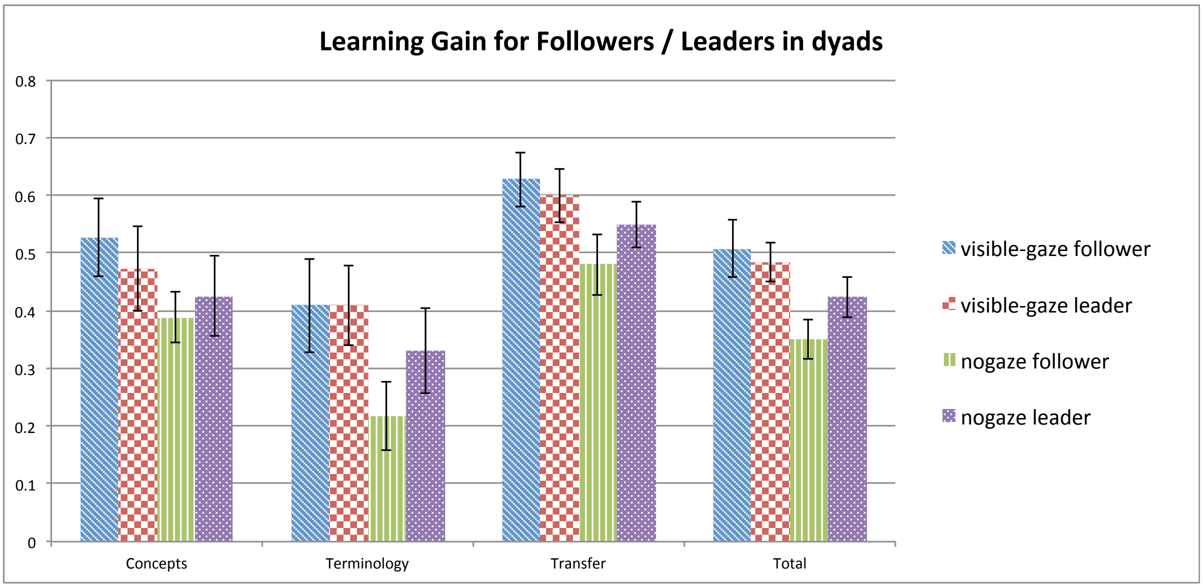

Figure 1. The total scores of the learning gain and the three sub-dimensions measured: conceptual understanding, participants' recall of the terminology, and transfer questions (crossed with two factors: experimental conditions and individuals' status in the dyad).

Eye-tracking Data

We isolated four kinds of measures from the eye-tracking data: first, we counted the number of fixations on the five contrasting cases and on the region showing the potential answers. Secondly, we aggregated the number of saccades between two regions from the six previously mentioned (i.e., 5 cases and 1 area for the answers). Thirdly, we defined a "joint attention" measure, where we counted how many times both participants looked at the same case on the screen. Previous research has shown that subjects need ~2 seconds to focus their attention on an object after a peer mentioned it (Richardson & Dale, 2005). We followed those guidelines to create our measure: for each data point, we checked whether the other member of the dyad was looking at the same area of the screen during the following two seconds. Fourthly, we used the size of the subjects' pupil as an indication of their cognitive load. Since eye-trackers react differently to different eyes' physiology, we divided each measure by the total number of data points for each subject. This yielded the percentage of fixations, percentage of saccades and percentage of joint attention. For cognitive load, we also subtracted the smallest value from each measure of a particular participant to take into account differences in eyes' morphologies. Participants' pupil size is not always a reliable measure, especially when the lighting conditions vary; however, since the room we used for the experiment did not have a window and thus had a constant lightning, we included those results for our analysis.

We excluded 5 subjects from those analyses because of missing data (i.e., the eye-tracker crashed during the activity). Three participants were in the "no-gaze" condition and two participants in the "visible-gaze" condition. We thus have 37 subjects when measuring the number of fixations and saccades and 16 dyads (32 subjects) when measuring joint attention. Due to space constraints, we will describe only a subset of our results.

We found that participants in the "no-gaze" condition had significantly more fixations on case 1 ($F(1,35) = 9.69$, $p < 0.01$), and case 3 ($F(1,35) = 4.92$, $p < 0.05$). Participants in the "visible-gaze" condition spent more time looking at the answers ($F(1,35) = 10.41$, $p < 0.01$). In terms of cognitive load, we did not find a significant difference between our two conditions: $F(1,35) = 1.09$, $p = 0.3$ (mean = 1.44, SD = 0.34 for "visible-gaze"; mean = 1.31 SD = 0.41 for "no-gaze"). The interaction between experimental condition and status in the dyad (i.e., leader or follower) is not significant: $F(1,29) = 2.51$, $p = 0.12$, but the effect size is between medium and large (partial eta squared = 0.08). It would be interesting to have more subjects to see if this result becomes significant. The pattern is similar to the one described for the learning test (i.e., followers have a higher cognitive load than leaders in the "no-gaze" condition, and a lower load than leaders in the "visible-gaze" condition).

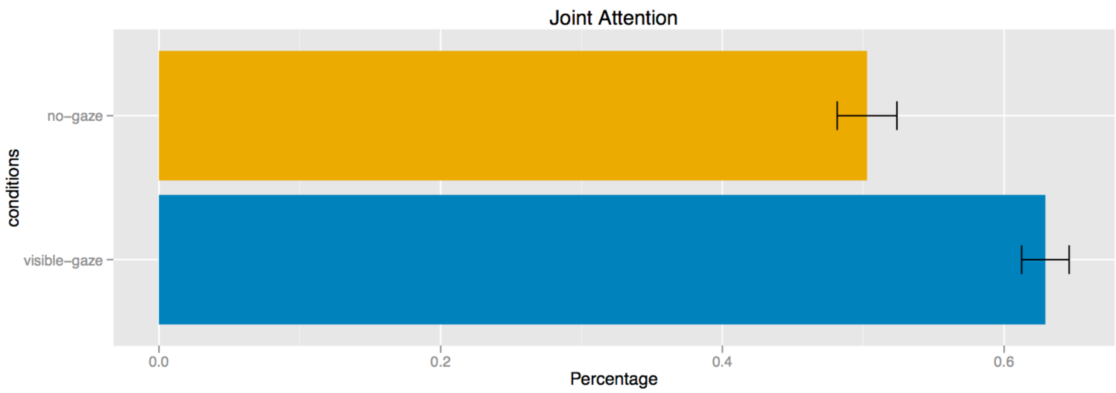

Figure 3. Percentage of Joint Attention in each experimental condition.

Participants in the "visible-gaze" condition achieved joint attention more often than the participants in the "no-gaze" condition (see Fig. 3): $F(1,30) = 22.45$, $p < 0.001$. This result holds when taking dyads (and not individuals) as the unit of analysis: $F(1, 14) = 16.36$, $p < 0.001$. The percentage of joint attention is one of the only measures correlated with a positive learning gain: $r = 0.39$, $p < 0.05$.

Model for Potential Mediators

In this section, we tested which process variables were most strongly associated with a positive learning gain. One may hypothesize that the quality of collaboration, the amount of effort produced by the participants, or the number of moments of joint attention may predict students' learning. We tested for multiple mediation using Preacher and Hayes' bootstrapping methodology for indirect effects (Preacher and Hayes, 2008). We used 5,000 bootstrap resamples to describe the confidence intervals of indirect effects in a manner that makes no assumptions about the distribution of the indirect effects. Significance is determined by checking if a confidence interval does not contain zero. We tested our model with the following candidates for being a mediator: collaboration, percentage of joint attention, cognitive load. GPA was used as covariate, since our goal is to find mediators regardless of participants' grades. Results for multiple mediation indicated that only joint attention (CI: [0.03; 0.19]) was a mediator for learning (see Fig. 4).

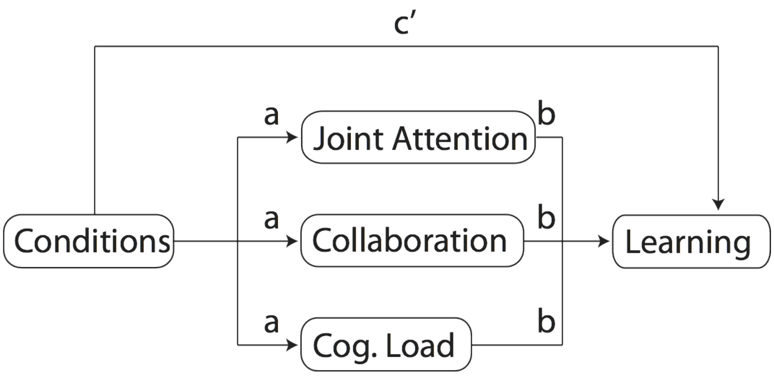

Figure 4. Mediation model for our data: we tested whether cognitive load (measured by the pupils' sizes), quality of collaboration (measured by Meier, Spada and Rummel's rating scheme) and percentage of joint attention (estimated with the eye-tracking data). Only joint attention was found to be a significant mediator.

Vignette

The previous section provides quantitative data on the effect of a gaze-awareness tool on students' remote collaboration. However it does not provide us with any explanation for causal mechanisms. Table 1 tries to suggest answers to this question by comparing two dyads in terms of their gaze patterns. We compared two groups: one in the "visible-gaze" condition (left side) and one in the "no-gaze" condition (right side). The main goal of this comparison is to illustrate how our intervention changed the behavior of our participants. More specifically, we focused on four dimensions: students' ability to coordinate themselves, create convention, build hypotheses and share theories.

<table>
<tr><th><u>V</u>isible-gaze (P54-P55)</th><th><u>N</u>o-gaze (P07-P08</th></tr>
<tr><td>
Lea (L) is the leader, and Flo the follower (F).

(v1) establishing common ground (0:51)
F: “I see this light blue dot”
L: “I think that’s my gaze”
F: “Oh cool... that’s so bizarre!”

(v2) building anticipation (1:22)
L: I have an answer for the... [gaze moving to case 1]
F: [gaze moving to case 1]
L: ... further most left one.
F: Okay. Where the lesion is the orange colored thing.

(v3) sharing hypotheses (3:45)
L: maybe lesion two is... [gaze moving back and forth between the two hemifields, eyes and optic nerves]
F: [her gaze is moving from Lea’s gaze to the other lesions]
L: those are both... they would be disrupted... I think that lesion two would be... [gaze moving to the second answer]
F: [gaze moving to the second answer]
L: the second one.
F: why do you say that?
L: because they are both going to be equal [gaze moving to the lesion where two optic nerves (one from the right hemifield, one from the left hemifield) are severed]
F: [gaze drifting to the same point] oh right.

(v4) creating implicit conventions (6:20)
F: let’s look at two again
[both gazes move to lesion 2]
L: everything is sort of cut off...
F: well it’s just the two in the middle
L: yeah so that would be... left-left [gaze moving to the left hemifield of the left eye, followed by Flo’s gaze] and right-right [gaze moving to the right hemifield of the right eye followed by Flo’s gaze]
F: [gaze moving to the second answer, followed by Lea’s gaze] ... which would be the second one.
L: yeah, which is the second one.

(v5) sharing theories (7:34)
L: so for the fifth we are not sure [...].
[both gazes are exploring different cases on their own]
L: so maybe the further away from the eye it is, the less severe [gaze moving from the eyes to the LGN on lesion 1]
F: [gaze moving to lesion 1] Maybe... what was lesion one again?
L: that was the top left and top right [gaze moving to the 4th answer, followed by Flo’s gaze], the fourth one down.
F: Oh... Ooooh... Hum [gaze comparing cases 1 and 5]
L: so the one you had was right by the eye, and it was completely crossed out [gaze on lesion 6]
F: [gaze moving to lesion 6, then 5] so maybe this would be similarly only a quarter of the eye
L: [gaze on answer 5] yeah, maybe it would be the third one from the bottom [followed by Flo’s gaze on answer 5]
F: maybe... hum... [gaze jumping from lesion 5 to answer 5]
</td><td>
Laurie (L) is the leader, and Fiona is the follower (F).

(n1) establishing common grounds (0 - 0:30)
L: Hi!
F: Hi! [laughing] I don't get this stuff.
L: I don't either!
F: okay, so I have one with the answer [looking at her case]
L: yeah I have an example too. [looking at her case]

(n2) sharing answers (2:37 - 3:45)
F: so... [gaze moving to lesion 1] do you see lesion 1?
L: yes [gaze moving to lesion 1].
F: I think it blocks Meyer's loop somehow.
L: yes
F: so the answer would be the left and the right...
L: [gaze moving moving between answer 4 and 5]
F: both the visions, they're blocked by one fourth. So it's no like completely blocked. So the answer would be that one.
L: but how is it... hum... [gaze still moving between answer 4 and 5]. So I think it's the fourth answer down? Where the quarter is blacked out on the top? On the left?
F: yes both right and left vision.

(n3) sharing theories (8:19)
F: [gaze on lesion 5] you said that lesion 5 would be the third from the bottom, right? [gaze on answer 5]
L: [gaze moving from lesion 5 to answer 5] yeah I think so because it's blocking the left lower part [gaze moving back and forth between lesion 1 and 5]
F: hu [gaze moving back and forth between lesion 4 and 6]
L: but then again it kinda doesn't make sense because if the answer for lesion 1 was the top left,
F:[gaze moving to lesion 1, but then going back to lesion 4] hu hu
L: then wouldn't it be blocked on the opposite side of where the lesion is?
F: [gaze moving from lesion 1 to 5] that's what I thought...

(n4) sharing hypotheses (5:20 - 6:10)
L: okay lesion 4...
F: lesion 4 would be
L: [gaze moving from the lesion to the third answer] I think it is the one that's half and half, the third one from the top. Because it blocks... [gaze moving to the eye]
F: [gaze moving to the third answer] the left part of the vision?
L: yeah I don't know
F: maybe
L: [laughing]
F: Hum... maybe. I don't know [laughing], whatever you say [both laugh]
</td></tr>
</table>

In terms of coordination, we found a strong difference between our two dyads. More specifically, the *sequence of actions* was reversed: in the “visible-gaze” dyad, the leader would start talking about a lesion, and the follower’s gaze would go to the same area on the screen *before* the leader mentioned the lesion’s number (v3). In the “no-gaze” dyad, the follower would have the double burden of finding the lesion of interest and following the leader’s explanation in parallel (n2). We argue that our intervention facilitated coordination, and helped the follower anticipate the leader’s explanations. Secondly, we found interesting conventions in the “visible-gaze” dyad (v4): when Lea says “that would be... left-left, right-right”, neither of them ever explicitly stated that she referred to the diagram’s eyes and hemifields. Rather, they implicitly build the convention of moving their gaze as a deictic gesture to complement their explanations. Thirdly, we hypothesize that our intervention helped students share their cognition, even though they did not master the expert terminology of the domain: sentences as vague as "they are both going to be equal" (v3) suddenly made sense when Lea pointed her gaze at the optic nerves to show that half of the information from each hemifield would be disrupted. This is particularly interesting because novices often lack the vocabulary to effectively communicate their assumptions. In our case, it provided Flo with additional information about the symmetry of the brain and helped her build her own hypotheses. Finally, we observed a tighter coupling between subjects' attention in the "visible-gaze" condition (v5): gazes would "dance" together during a longer period of time and focus on the same lesions even though they were not explicitly mentioned. In the control group, the follower would briefly attend to the same lesion as the leader and then continue to explore other lesions (n3). This suggests that the theories built during the activity were more the results of the dyad's shared cognition (in the "visible-gaze" condition), and more the results of individuals' contribution in the "no-gaze" condition.

Discussion

Our findings demonstrate the importance of supporting joint attention in collaborative learning activities. We conducted a study where students needed to learn from five contrasting cases in a remote collaboration. In one condition, subjects could see the gaze of their partner on the screen as it was being produced. In the other, they could not. Our results reveal that this simple intervention was associated with subjects in the first group producing a higher quality of collaboration and learning more from the contrasting cases. In particular, subjects characterized as *followers* saw their learning gain dramatically increase. This result was partially confirmed by a similar pattern found for students' cognitive load: followers in the control group spent more effort than leaders while learning less; followers in the treatment group spent less effort than leaders but learned more. We also found that subjects in the "no-gaze" condition spent more time on cases 1 and 3; this suggests that they took more time (and probably had more difficulty) sharing their answers. Participants in the "visible-gaze" condition had a higher percentage of joint attention, which proved to be a significant mediator for learning.

These results provide strong evidence for the important contributions of **real-time mutual gaze perception**—a special form of technology-mediated shared attention—to the learning gains and collaboration quality of collaborative learning groups. Additional qualitative observations suggest that our intervention helped students on four dimensions: by supporting coordination, creating conventions, sharing cognition and by making knowledge-building a collective process rather than an individual one.

One might argue that a shared pointer could achieve a similar effect. We believe that real-time mutual gaze perception has several key advantages over a shared pointer: first, there is a cognitive overhead associated with consciously moving a cursor to a region of interest, which may interfere with the learning task. A gaze awareness tool does this work automatically, without requiring additional effort from the user. There is also a certain amount of uncertainty associated with a cursor that stopped moving; is your partner thinking, being distracted, or waiting for you? By looking at the videos of our experiment, we saw that members of a dyad would perform some sort of "micro-monitoring" of their partner's behavior, where they would check on their partner's gaze every few seconds. We believe that a continual flux of gaze information reduces uncertainty and helps students regulate the dynamics of their dyad. In summary, we hypothesize that our gaze awareness tool enabled some behaviors that would not be possible with a shared pointer. Future studies are needed to demonstrate the unique affordances of each of those interventions.

This study has limitations. First, we studied a very specific kind of collaboration: situations where members of a dyad were communicating via a microphone and sharing a computer screen. It is not clear whether this kind of awareness tool would have the same effect in a co-located situation; one may assume that joint attention is easily achieved a in face-to-face or side-by-side collaboration, but key papers in the learning sciences suggest that it may not be the case (e.g. Barron, 2003). Future studies using eye-tracking goggles on interactive surfaces will answer this question. Secondly, students had a very limited amount of time to work on the contrasting cases. It is unclear how this limitation impacted students' performance. Thirdly, we only cursorily evaluated the transcripts of the dyads. More fine-grained coding schemes would provide additional clues as to how joint attention facilitated collaborative learning; the interaction effect between followers and leaders is especially interesting and should be analyzed in greater depth. Lastly, one may argue about the sub-categories describing the learning gains (e.g. it is debatable whether the questions about predicting the effect of a lesion are effectively measuring conceptual understanding); however, because we are not making particular claims about those sub-categories, and since the same pattern is repeated across our three learning sub-dimensions (i.e., the interaction effect between followers and leaders), we do not consider this issue to be a serious limitation of our findings.

In future work, we plan on evaluating the result of our qualitative observations. More specifically, we want to quantitatively measure the four dimensions we uncovered and show that those processes are significantly different across conditions. Secondly, a next logical step is to investigate this phenomenon in a more natural setting (e.g., in a co-located situation). Eye-tracking goggles could offer an interesting tool for this purpose. Thirdly, it would be interesting to see if those results generalize beyond contrasting cases; it may be that this intervention is only effective for perceptual tasks. Finally, our results suggest that supporting joint attention between novices and experts would bring interesting results, as real-time mutual gaze perception provides a form of "inter-identity technology" (Lindgren & Pea, 2012). As followers, novices could more easily share their understanding of concepts without having to know the expert terminology; additionally, it would disambiguate experts' explanations by providing perceptual clues to novices (Hanna & Brennan, 2007).

Endnotes

(1) Attentional alignment is also established partly by body position and orientation (Kendon, 1990).
(2) The text used in the second part of the study is accessible here: http://www.scribd.com/doc/98921800. Originally retrieved from Washington University in St-Louis (http://thalamus.wustl.edu/)

References

Aronson, E., Blaney, N., Sikes, J., Stephan, G., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage Publication.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*(3), 307-359.

Bates, E., Thal, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology, 25*, 1004-1019.

Cherubini, M., Nüssli, M., and Dillenbourg, P. 2008. Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. *In Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia, March 26 - 28, 2008). ETRA '08. ACM, New York, NY, 173-180.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Hanna, J. E. & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*, 596-615.

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, *1*(1), 77–89.

Jermann, P., Mullins, D., Nuessli, M.-A., Dillenbourg, P. (2001). Collaborative Gaze Footprints: Correlates of Interaction Quality. *In Proceedings of CSCL 2011.* Hong Kong.

Kendon, A. (1990). *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

Leudar, I., Costall, A., & Francis, D. (2004). Theory of Mind A Critical Assessment. *Theory & Psychology*, *14*(5), 571–578.

Lindgren, R., & Pea, R. (2012, July). Inter-identity technologies for learning. *Proceedings of the International Conference of the Learning Sciences (ICLS2012: The Future of Learning)*. Sydney, Australia.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer- Supported Collaborative Learning, 2*, 63–86.

Preacher, K. J., and Hayes, A. F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods 40*, 3 (2008), 879.

Richardson, D. C., & Dale, R. (2005). Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*, 1045–1060.

Richardson, D.C., Dale, R., & Kirkham, N.Z. (2007). The Art of Conversation Is Coordination: Common Ground and the Coupling of Eye Movements During Dialogue. *Psychological Science, 18*(5), 407-413.

Roth, W. M. (2001). Gestures: Their role in teaching and learning. *Review of Educational Research, 71,* 365–392.

Sangin (2009). *Peer knowledge modeling in computer supported collaborative learning*. (Doctoral Dissertation). Retrieved from http://infoscience.epfl.ch/record/133637.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences, 4*, 321-354.

Schwartz, D. L. & Bransford, J. D. (1998). A time for telling. *Cognition & Instruction*, 16, 475-522.

Stem, D. (1977). *The first relationship*. Cambridge, MA: Harvard University Press.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.