

Real-time mutual gaze perception enhances collaborative learning and collaboration quality

4
5

Bertrand Schneider · Roy Pea

6

Received: 7 May 2013 / Accepted: 17 September 2013

7

© International Society of the Learning Sciences, Inc. and Springer Science+Business Media New York 2013

8

Abstract In this paper we present the results of an eye-tracking study on collaborative problem-solving dyads. Dyads remotely collaborated to learn from contrasting cases involving basic concepts about how the human brain processes visual information. In one condition, dyads saw the eye gazes of their partner on the screen; in a control group, they did not have access to this information. Results indicated 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.

9
10
11
12
13
14
15
16**Keywords** Collaborative learning · Awareness tool · Eye-tracking17
18

Introduction

19

Foundational work in developmental psychology and in the learning sciences demonstrates that joint attention plays a crucial role in any kind of social interaction: From babies learning from their caregivers to parents educating their children, teenagers learning from school teachers, students collaborating on a project or for any group of adults working toward a common goal, joint attention is a fundamental mechanism for establishing common ground between individuals. Our goal is to design technological interventions to facilitate this process.

20
21
22
23
24
25

Technically, 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 such as pointing at a focus of interest

26
27
28
29
30
31
32
33
34

B. Schneider (✉) · R. Pea
Graduate School of Education, Stanford University, Stanford, CA, USA
e-mail: schneibe@stanford.edu

R. Pea
e-mail: roypea@stanford.edu

to establish joint visual attention so as to signal important features of the environment to their children (Bates et al. 1989). Professors and mentors teach by highlighting subtle nuances between the conceptual understanding of their students and experts (Roth 2001). Groups of students manage 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 joint attention rests significantly though not entirely¹ 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). Those two mechanisms are subject to inefficiencies because misunderstanding can happen on a verbal and on a physical level. 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 patterns that separate them from experts. For instance, Biderman and Shiffrar (1987) showed that experts in performing chick-sexing can categorize 1,000 chicks per hour with an accuracy of 98 %, but those experts have a lot of trouble explaining to novices (researchers, in that case) how they reached such an impressive speed and precision. Thus, words are sometimes a clumsy medium for teaching perceptual skills. Physically, there is an extra step of taking the point of view of the other 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 et al. 2004).

The goal of our work is to develop new ways of supporting the establishment of perceptual joint attention (as distinguished from cognitive, or social joint attention). Our assumption is that higher levels of visual synchronization are positively associated with students' quality of collaboration and learning experience. In our study, we designed an intervention to increase the quantity of students' number of moments of joint attention and studied the effects of the intervention on several interlaced variables: visual synchronization, quality of collaboration and learning gains computed from pre and post-test. We use eye-tracking technologies to make it possible to share users' real-time gaze behaviors during collaborative learning. More specifically, our first attempt in this study involved dyads in a remote collaboration studying contrasting cases (Schwartz and Bransford 1998). We introduce a new kind of awareness tool that provides participants with the continuous updating of the position of their partner's gaze on the screen. Thus, we depict our intervention as enabling *real-time mutual gaze perception*.

In the following section, we describe previous research in studying joint attention in collaborative learning situations. We then survey studies using eye-trackers and previous attempts at developing "awareness tools" in CSCL (i.e., tools that provide additional information to students about their peers). We conclude by summarizing the literature on joint attention and by formulating our research questions.

Previous work on joint attention and awareness tools

Developmental psychologists have conducted the vast majority of the work on joint attention by highlighting the crucial role of gaze coordination between infants and adults during language learning. Since this work is a primary inspiration for our research, we will start by briefly describing a few foundational studies in this area of inquiry and conclude by sketching the significance of those results for the field of the learning sciences.

¹ Attentional alignment is also established partly by body position and orientation (Kendon 1990).

The fundamental role of joint attention in infancy

78

An important developmental milestone is the ability to coordinate visual attention between a partner and an object of interest. Several studies suggest that humans acquire this skill early in life. Baldwin (1981) showed that 16-month-old babies are able to detect nonverbal cues to a referenced object. Bakerman and Adamson (1984) demonstrated that with age, person engagement (i.e., “the infant is engaged just with the other person. Typically such engagement involves face-to-face or person play; For example, a baby giggles and coos as his mother places her face close to his and tickles him”) declined while coordinated joint engagement increased (i.e., “The infant is actively involved with and coordinates his or her attention to both another person and the object that person is involved with. For example, the baby pushes the truck the mother has been pushing and then looks back and forth between the mother’s face and the truck”); additionally, an infant’s social coordination was more likely to happen when the child played with his or her mother. Charman et al. (2000) followed 13 infants aged 20 months for 2 years and administered a battery of cognitive tests at different intervals; across a variety of different measures, they found that only joint attention behaviors were longitudinally associated with increased theory of mind abilities 2 years later. Those studies showed that attentional deployment is one of the first social and emotional regulatory processes to appear. Indeed, young infants communicate their emotions by being in a state of visual synchronization with their caregivers—which in turn help them achieve visual coordination when learning to speak (Stern 1977). Without the ability of establishing joint attention, infants would have much more trouble acquiring their native language exemplified by the studies of autistic children who show impoverished joint attention behaviors (Mundy et al. 1990), and by studies indicating how greater gaze following by infants in play sessions with their mothers predicts faster vocabulary development (Brooks and Meltzoff 2008).

For the scope of this paper, we will not conduct an exhaustive review of the developmental work in this field. However, we can confidently assume that joint attention is an established and relevant concept in developmental and social psychology: meaningful interactions have been shown to be associated with repeated moments of joint visual attention. Humans need to make sure that they are communicating about the same object of interest to avoid misunderstanding. The previous paragraphs demonstrated that babies learn language, in part, by establishing visual coordination with their parents, and that higher levels of joint attention facilitate language acquisition. The following paragraphs suggest that children and teenagers also learn more efficiently by being visually synchronized with their peers.

Joint attention in the learning sciences

110

During the past decades, research in education has focused substantial efforts on social learning and small group cognition. The inspiration for this effort mainly comes from Piaget (1928), who postulated that socio-cognitive conflicts cause major cognitive restructuring, and Vygotsky (1978) who claimed that learning happens first on a social or cultural level, which is then internalized. Those two theories have been joined together under the umbrella of *socio-constructivist* theories of learning. This approach emphasizes the importance of collaboration and negotiation of meaning for thinking and learning; as a consequence, socio-constructivist researchers have devoted their attention to analyzing group interaction and identifying characteristics of successful patterns of collaboration. Over the past two decades, CSCL researchers have begun to extensively study the influence of technology on collaborative learning. A good summary of the goals of the CSCL field can be found in Dillenbourg et al. (1996, pp. 189–211). And we may note that joint attention is associated with many overlapping concepts in the learning sciences and CSCL—“shared cognition,” “intersubjectivity,” “grounding processes in

conversation,” “joint problem-solving,” and “distributed cognition” (please refer to Barron and Roschelle 2009, for more details about these overlapping concepts).

However, as Salomon and Globerson (1989) point out, teams do not always function the way that they should. There are multiple issues that can arise in collaborative learning situations (e.g., the “Free Rider effect”, referring to those who benefit from the collaborative activities of the group without contributing their own efforts, or the “Sucker Effect”, a tendency for participants to contribute less to a group if they expect others will think negatively of them if they work too hard or contribute too much). Group work can lead to unproductiveness, wasted time and feelings of discouragement. More specifically, Barron (2003) begins to unpack the complexities of collaborative work with detailed analyses of triads solving mathematics problems. Focusing on explanations for variability in outcomes, she contrasted two groups of students who produced radically different outcomes; in one group, students generated, confirmed, documented, and reflected upon correct proposals. In the other group, students generated correct proposals but their partners ignored or rejected them without rationale and left them undocumented. Case-based portraits depicted the challenges that arose as participants attempted (or did not) to coordinate individual perspectives into a joint problem solving space. In the less successful case, relational issues arose that prevented the group from capitalizing on the insights that fellow members had generated. Such relational issues included competitive interactions, differential efforts to collaborate, and self-focused problem-solving trajectories. Behaviorally, these issues were manifest in violation of turn-taking norms, difficulties in gaining the floor, domination of the group workbook, and competing claims of competence. Those differences were not explained by students’ prior achievement; rather, the mutuality of exchanges and the achievement of joint attention were found to be better predictors of the groups’ success. It seems that the outcome of collaboration not only depends on individuals’ contributions, but also on how well group members manage individual and joint attention during the collaborative activities.

As a consequence, we will adopt the point of view expressed by Dillenbourg et al. (1996), who argued that: “collaboration is in itself neither efficient nor inefficient. Collaboration works under some conditions, and it is the aim of research to determine the conditions under which collaborative learning is efficient”. Our goal goes beyond observing collaboration: we are interested in designing technological interventions that will support and increase the quality of collaboration. This goal is shared among many researchers in CSCL. More specifically, we base our intervention on the findings of Barron’s (2003) study: If joint attention was among the strongest predictors of a good collaboration, then facilitating this process should lead to more productive social interactions.

Awareness features in CSCL

As mentioned above, teams are not always more efficient than individuals: group members need to sustain mutual understanding, manage a smooth flow of communication, gather as many solution-relevant pieces of information as possible, reach a consensus, divide tasks equally, make sure to finish the current task within the time limit, treat each other with respect, and actively engage in finding a relevant solution to the problem at hand (Meier et al. 2007). With so many constraints, it should not be surprising that a good collaboration is difficult to establish and maintain. One promising approach in supporting group collaboration has emerged in CSCL over the last decade: Researchers have begun to design *awareness tools* to support productive interactions among students. Awareness tools provide additional information to a group of students about their peers (e.g., their level of expertise, extraversion, or progression toward a goal). Multiple studies have found that awareness tools increase the quality of collaboration in small teams.

For instance, Sangin (2009) studied pairs of students remotely working on a concept map and found evidence that a knowledge awareness tool (i.e., displaying the level of expertise of each member of the dyad) was associated with a higher density of gaze-coupling to a joint referent (i.e., joint attention), a higher quality of collaboration and increased learning gains. In another line of work, Bachour et al. (2010) described the design of an interactive table displaying the participants' level of participation (i.e., as indexed by the amount of speech produced by each individual); their empirical evaluation suggests that this simple visualization leads to more balanced patterns of collaboration. More specifically, it prevented those who might be described as extroverted users from dominating the discussion and discouraged under-participation from those who might be described as introverted individuals². Independently but with convergent results, Kim and Pentland (2009) used sociometric sensors to detect group dynamics and found that when these data were used to provide real-time feedback to participants, speaking time and interactivity level of groups changed significantly. Especially interesting was that in groups with one dominant person, the feedback effectively reduced the dynamical difference between co-located and distributed collaboration as well as the behavioral difference between dominant and non-dominant individuals. Finally, a slightly different kind of awareness tool supporting interactions between teachers and students was described by Alavi and Dillenbourg (2012). They built an ambient awareness tool to support collaborative work in recitation sections. Each device looks like a lantern and displays the status of the group (e.g., which exercise students are working on, if they have asked for help, how much time they have been waiting). A user study suggested that this kind of tool leads to improved interactions between teams and tutors; they wasted less time waiting for the teaching assistants and spent more time working on their assignments. This last example is conceptually different from the other projects, because it provides awareness at a higher level of social organization (i.e., between a group of students and a teacher). It is slightly less relevant to us since we are interested in raising students' awareness of each other's learning activities as indicated by gaze patterns to learning resources and simultaneous audio channel interchanges.

These four projects show that simple visualizations can be quite powerful for supporting interactions in small groups. As a consequence, we propose to build on this promising body of work to help students coordinate joint visual attention with eye-tracking technologies. In the following section, we summarize existing work on using eye-trackers in education and describe our approach in designing an awareness tool for supporting visual coordination.

Eye-tracking and joint attention

Even though the first eye-trackers were built and used in research over a century ago (e.g., Dodge and Cline 1901; see Jacob and Karn 2003), their use is not widespread in the scientific community. Costs, technological challenges, accuracy and latency, the need for advanced data analysis skills and other obstacles have prevented their propagation. However, the ability to track subjects' gaze can provide rich and insightful data; some researchers even reflect on how eye-trackers may open a new "window into the mind" of the users (Duchowski 2007) since visual attention often reflects cognitive processes. On a technical level, eye-tracking devices generate three kinds of data: *saccades* ("jumps", that reposition the fovea on a new location of the visual field), *fixations* (prolonged focus of attention on a specific location) and *smooth pursuits* (following an object on the screen). Combined together, these measures provide unique opportunities to understand people's cognitive processes. Furthermore, several eye-tracking

² It should be noted that this study did not employ empirical measures of extroversion or introversion to arrive at these characterizations.

devices used in parallel may afford an indication of the level of synchronization of the different members of a group; for instance, by measuring the number of times users look at the same area on the screen within a specified time window (i.e., number of moments of joint attention).

Previous work in CSCL used eye-trackers to study joint attention in collaborative learning situations. For instance, Richardson and Dale (2005) found that the degree of gaze recurrence between individual speaker—listener dyads (i.e., the proportion of times that their gazes are aligned) is correlated with the listeners' accuracy on comprehension questions. Richardson et al. (2007) showed that common knowledge grounding (i.e., hearing the same background information before the task) positively influenced the coordination of visual attention in a spontaneous dialogue. Jermann et al. (2001) used synchronized eye-trackers to assess how programmers collaboratively worked on a segment of code; they contrasted a 'good' and a 'bad' dyad, and their results suggested that a productive collaboration is associated with high joint visual recurrence. In another study, Nuessli (2009) showed that eye-tracking data can be integrated with other measures to build models of group behavior: by using gaze and raw speech data (pitch and speed of the voice), he was able to predict participants' success with an accuracy rate of up to 91 %. As importantly, he was able to make this prediction before the activity was over. In a similar study, Liu (2009) used machine-learning techniques to examine gaze patterns for collaborating dyads, and was able to predict the level of expertise of each subject as soon as 1 minute into the collaboration (with 96 % accuracy). In a similar way, Cherubini et al. (2008) designed an algorithm for detecting misunderstanding in a remote collaboration by using the distance between the gaze of the emitter and the receiver; they found that if there is more dispersion, the likelihood of misunderstandings is increased. Finally, Brennan et al. (2008) studied the effect of shared gaze and speech during a spatial search task; they found that the shared gaze condition was the best of all. It was twice as fast and efficient as solitary search, and significantly faster than other collaborative conditions.

Taken together, those results support the idea that joint attention and, more generally, synchronization between individuals, is crucial for an effective collaboration. They also suggest that eye-trackers are a promising way to understand and influence the factors responsible for a high-quality collaboration.

Summary of previous work and hypotheses

Based on prior work studying joint attention and the effects of awareness tools on collaborative learning, we conjecture that new technologies can facilitate collaboration by supporting the establishment of joint attention. *In a unique application of eye-tracking technologies, we propose that their use to inform a collaborator about their partner's gaze during a collaborative learning situation by creating a new real-time perceptual data stream overlaid on the static representation of the learning resource that they each are studying.* We go beyond prior research using eye tracking as a researcher methodology and representational medium for making scientific inferences about learners or collaborating learners, to use eye tracking in order to provide a new real-time information resource for learners to exploit for enhancing their own collaborative processes.

More specifically, our first attempt in this vector of innovation involves dyads studying contrasting cases (Schwartz and Bransford 1998) in the domain of neuroscience. In our study, contrasting cases were designed "to help students notice information they might otherwise overlook. As with tasting wines side by side, contrasts can improve discernment" (Schwartz and Martin 2004). We followed the examples given in Schwartz and Bransford (1998) to help students notice the deep structure of the concepts taught. More specifically, this learning activity is based on the "Preparing for Future Learning" (PFL) framework. The PFL framework

proposes to design perceptual tasks to prepare students for traditional instructional activities (e.g., attending a lecture or reading a textbook chapter). The PFL approach encourages students to explore in order to generate their own theories about a class of phenomena, which sets the stage for future learning. Note that we are not testing the PFL approach *per se*, but we decided to use it for three reasons: First, we assumed that a gaze awareness tool is more likely to produce positive outcomes for a perceptual task, since we are enhancing the ways in which students may perceive their peers' visual behavior. Second, we are interested in improving proven pedagogical strategies; the PFL framework is recognized as being a fruitful approach for teaching in complex domains. For these reasons, our work is more likely to have an impact on existing classroom practices when eye-trackers become commonly used in everyday life. Finally, we care about supporting students' transfer of knowledge to new situations (Pea 1987), as opposed to rote memorization. PFL activities are known to promote higher gain on transfer questions (Schwartz and Martin 2004).

General description of the experiment

Our experiment had three distinct steps: during the first 12 min, dyads worked on five contrasting cases in neuroscience that were represented in a single static diagram. We choose contrasting cases as an instructional approach, because joint attention is more likely to be a significant mediator for students' learning gains in a highly perceptual task. In this experiment, we were specifically targeting deictic behavior: by providing gaze information from the participants' partner, we eliminated the need for them to precisely describe which area of the screen they were referring to. However, as we note in our discussion, it is unclear if the results we obtain will generalize beyond diagram-based contrasting cases. Students had to collaboratively explain how visual information is processed in the human brain by studying the models described in Fig. 1. In the second step, they then read a text on the same topic for 12 min. In the final step, they answered a learning test. We used a between-subjects design with two conditions. In one condition ("visible-gaze"), dyads were able to see the gaze of their partner on the screen. In the other condition ("no-gaze"), they could not. In both conditions, an audio channel was open between the collaborating participants.

Our hypotheses for results from the two conditions are as follows: first, we expect the dyads in the treatment group (i.e., students who could see their partner's gaze on the screen) to have a higher quality of collaboration, since this visualization will disambiguate their focus of attention and better enable "common ground" for learning conversations (Clark and Brennan 1991). Second, we assume that a better collaboration will be positively associated with participants' learning gain (Barron 2003), since users will more efficiently communicate their understanding of the content taught, and thus better explore the problem space for the information needed for their learning task.

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 ($N=22$) with 15 females and 7 males; the control group was in the "no-gaze" condition, with 13 females and 7 males ($N=20$). There was no significant difference in terms of GPA (Grade Point Average) 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

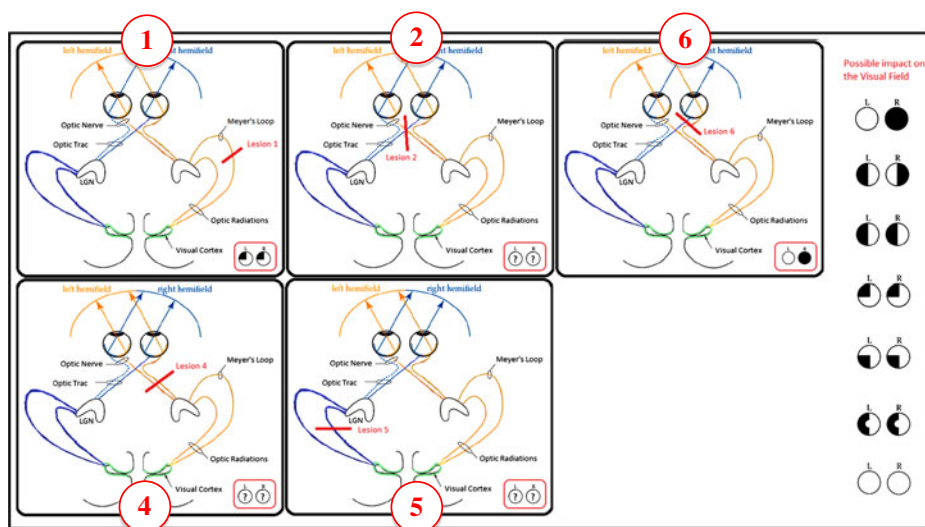


Fig. 1 The dyads worked on the five contrasting cases above. Possible answers are shown on the right side. Answers of two cases (#1, top left and #6, top right) were given to subjects. Participants had to solve the three remaining cases (#2—top middle—and #4 and #5—bottom left and right of the screen,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

psychology and were required to participate in an experiment as part of their course. No participant had previous knowledge in neuroscience before completing the task. Participants did not know each other prior to the study.

Material During the first step of the experiment, dyads worked on the contrasting cases shown in Fig. 1. Their task was to infer the effect of three particular lesions (labeled 2, 4 and 5 in Fig. 1) on the visual field of a patient. Students had two main ideas to discover to be successful in this learning task: first, visual information is crossed after the left geniculate nucleus (LGN): in general, the left hemisphere of one's brain processes the information coming from one's right side of the visual field, and the right hemisphere of one's brain processes the information coming from the left side of one's visual field. Second, participants had to discover that visual information is again divided between the LGN and the visual cortex: the outer optic radiation (called Meyer's Loop) processes information coming from the top half of the visual field, while the inner optic radiation processes information coming from the bottom half of the visual field. Thus, each pathway between the LGN and the visual cortex carries information relative to a quarter of one's visual field. To derive the correct answers, students needed to look both at the color coding of the contrasting cases and the answers for cases 1 and 6.

To compel learner collaboration, the answer for lesion 1 (top left) was visible only to the first member of the dyad while the answer for lesion 6 (top right) was shown only to its second member. This "jigsaw" method is commonly used to make sure 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 online³. The document was 5 pages long, contained 972 words and included 6 large figures. The content

³ The text used in the second part of the study is accessible here: <http://www.scribd.com/doc/98921800> (last access: 03/08/2013). Originally retrieved from Washington University in St-Louis (<http://thalamus.wustl.edu/>).

focused on the visual pathways shown on the contrasting cases in Fig. 1. It explained why various lesions have different impacts on the visual field. We removed unnecessary paragraphs with heavy medical terminology to keep the task doable in the amount of time provided to the students.

Experimental design Our study used a between-subjects design. Participants were randomly distributed between two conditions: In the treatment group (“visible-gaze”), dyads were able to see the gaze of their partner on the screen. In the control group (“no-gaze”), they could not. The gaze was only visible during the first step of the experiment for the treatment group (i.e., when participants had to collaboratively solve the contrasting cases shown in Fig. 1).

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 audio channel. The experimenter explained that participants would learn basic concepts in neuroscience, and he described the structure of the experiment—12 min of analysis of contrasting cases, 12 min of reading a text solitarily and silently, 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. At the beginning of the task, the contrasting cases were then presented for approximately 1 minute to each participant and the experimenter ensured that they understood the goal of the task. Participants then worked on the contrasting cases and tried to determine how different lesions affected the brain’s visual field. After 12 min, the screen automatically switched to a text explaining how the human brain processes visual information. The experimenter informed participants that they should read the text individually and afterwards discuss it with their partner. The audio channel remained opened for this step. After 12 min, the screen being observed by the dyads 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.

Eye-tracking setup We used two desktop-based Tobii X1 eye-trackers running at 30 Hz to capture and display participants’ gaze. We used an in-house server for synchronizing the two devices. Calibrations were performed using a five-points calibration at the beginning of the experiment. Additionally, since our areas of interest are quite large (for most analyses we used one diagram as an area of interest), we did not correct for gaze deviations. However, we cursorily watched the videos showing the participants’ gaze patterns to ensure that no large deviation was present in our dataset.

Design of the gaze-awareness tool The gaze of each participant was displayed to their partner as a light blue dot of 20 pixels of diameter on the screen. The circle was half transparent (40 % opacity) and refreshed approximately four times per second. We determined those values by trial and error to avoid a distracting effect; we found non-transparent circles that were refreshed too often created frustration. During the experiment, students had the opportunity to hide the circle by pressing any key on the keyboard, yet no participant used this function.

Measures Because no participant had previous knowledge in neuroscience, learning gains were computed from the final learning test, which contained 15 questions: five terminology questions (participants were asked to provide the name of a specific brain region or pathway), five conceptual questions (participants had to predict the effect of a specific lesion), and five

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?”). The tests were administered electronically and multiple-choice questions were coded automatically (i.e., students were allowed to choose only one option). Transfer questions were open-ended; we gave 1 point for a correct answer, 0.5 points for an ambiguous answer suggesting a correct logic, and 0 points for wrong answers. Two researchers evaluated the answers and agreed on a common definition for “ambiguous”, “correct” and “wrong”.

The quality of collaboration was rated using dimensions developed in Meier et al. (2007), who assessed collaboration on a five-point scale across nine 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 rendered it as an appropriate tool for assessing collaboration. In addition to those nine dimensions, we also computed a general “collaboration score” by averaging each dyads’ scores on those sub-dimensions. This compound variable allowed us to run higher-level correlations between our eye-tracking indicators and students’ interactions.

Additionally, we categorized each participant in a binary manner as being either the “leader” or the “follower” in the activity. This distinction is motivated by recent work in HCI (Human-Computer Interaction), where Shaer, et al. (2011) noticed that pairs of participants tended to assign “roles” to their members; for instance, in collaborative tasks there tends to be a “driver”, who is physically active and controls the interface, and a “passenger”, who is physically inactive and merely proposes verbal suggestions. This pattern was also documented in classroom-based uses of microcomputers in Sheingold et al. (1984), in which the pattern of “I’m the thinkist,

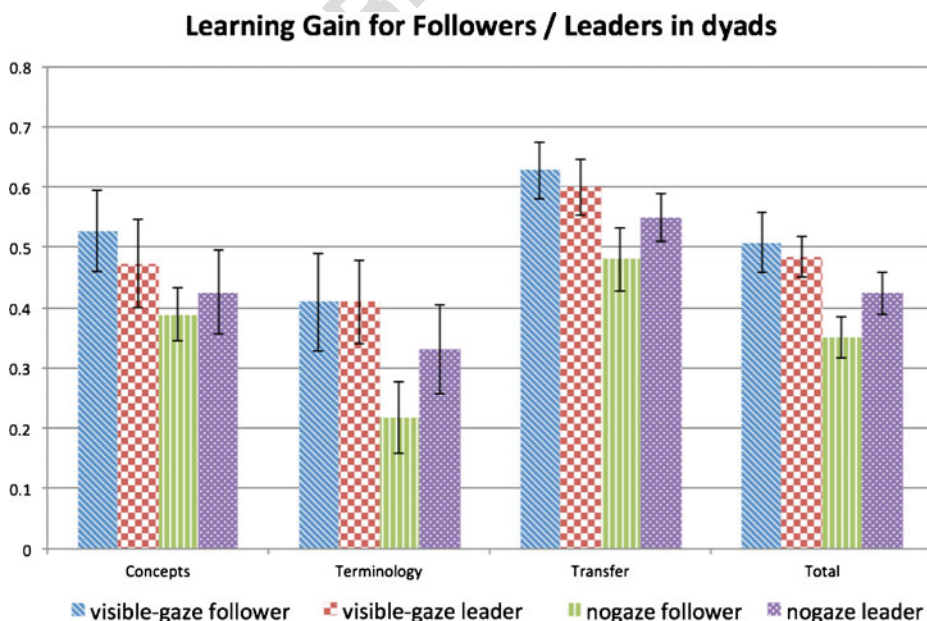


Fig. 2 The total scores of the learning gain and the three sub-dimensions measured: conceptual understanding, participants’ recall of the terminology, and transfer questions (crossed with two factors: experimental conditions and individuals’ status in the dyad). Whiskers represent standard errors

you're the typist" was identified. Those profiles of collaboration are interesting, because they inform us about the emergent dynamic of a group. Inspired by this approach, we developed a rough coding scheme to distinguish between 'leaders' (the equivalent of Shaer's 'drivers') and 'followers' (the equivalent of Shaer's 'passengers'). We used several indicators to categorize each dyad's members: 1) who started the discussion when the experimenter leaves, 2) who spoke most, 3) who managed turn-taking (e.g., by asking "what do you 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"). This measure can be considered as an aggregate estimation over the whole activity of the dyad's dynamic profile, since we acknowledged that subjects are likely to shift roles while solving contrasting cases. We also recognize that this categorization is more likely to be a continuum, and that in a few cases the difference between followers and leaders may be subtle. The decision to only have two categories was made to simplify the coding process and present clearer results (Fig. 2), but we acknowledge that this coding presents an oversimplified picture of the dyads' dynamic. Future work will analyze this kind of interaction on a more fine-grained level.

Finally, we 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 as a measure of cognitive load, as we explain subsequently. *Our main motivation for collecting and analyzing eye-tracking data is to compute a quantitative measure of joint attention.* By synchronizing our two eye-trackers, we can see how often dyads looked at the same thing at the same time on the screen. We can then relate this measure to the groups' quality of collaboration and learning gains. Additionally, we can also track students' cognitive load and see: 1) if monitoring an additional channel of information (i.e., the gaze of their partner) increased participants' efforts to complete the task, and 2) if more successful students are characterized by a higher level of cognitive load.

More specifically, from our data we isolated four different measures from the eye-tracking data:

- (1) First, we counted the number of fixations on the five contrasting cases and on the region showing the potential answers;
- (2) Second, we aggregated the number of saccades between two regions from the six previously mentioned (i.e., five cases and one area for the answers);
- (3) Third, 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 s to focus their attention on an object after a peer mentions it (Richardson and 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 preceding or following two seconds.
- (4) Fourth, we used the size of the participants' pupil as an indication of his or her cognitive load. When a person is faced with a challenging cognitive task, his or her pupils dilate (the task-evoked pupillary response: Beatty 1982; Beatty and Lucero-Wagoner 2000), so pupil dilation may be used for estimating cognitive load. However, it should be noted that there is some debate about using pupil dilation as a measure of cognitive load; consequently, our data for pupil dilation should be provisionally taken as estimations of cognitive load. Additionally, since eye-trackers react differently to the physiology of different eyes, we divided each measure by the total number of data points for each participant. This computation yielded the percentage of fixations, percentage of saccades, and percentage of joint attention. For the cognitive load, we also subtracted the smallest value from each

measure of a particular participant to take into account differences in eyes' morphologies. 443
 Participants' pupil size is not always a reliable measure, especially when the lighting 444
 conditions vary; however, since the room we used for the experiment did not have a 445
 window and thus had a constant lighting, we included those results for our analysis. 446

Qualitative analysis The previous measures provide quantitative data on the effects of a 447
 gaze-awareness tool on students' remote collaboration. However they do not provide us with 448
 any explanation for causal mechanisms that may be responsible for such effects. We also 449
 tried to qualitatively analyze our data by comparing two dyads in terms of their gaze 450
 patterns. We compared two groups: one in the "visible-gaze" condition and one in the 451
 "no-gaze" condition. The main goal of this comparison is to illustrate how our intervention 452
 changed the behaviors of our participants. More specifically, we focused on four dimen- 453
 sions: (1) students' ability to coordinate themselves, (2) to create convention, (3) to build 454
 hypotheses and (4) to share theories. We chose those two groups randomly; it is possible that 455
 there are differences between them that go beyond their experimental condition, but our goal 456
 here is not to generalize our observations to the entire sample or to a population. Rather, as 457
 stated above, our aim was to suggest potential mechanisms for the effect of a gaze-awareness 458
 tool on students' collaboration. Thus, we only watched the two videos at 0.5x speed to be 459
 able to analyze gaze patterns. We report our observations in the last subsection of the results. 460
 461

Results 462

In this section, we compare main effects for learning gains and collaboration scores across 463
 our two experimental groups. We then characterize the dyads of our experiments in terms of 464
 their gaze patterns by analyzing our eye-tracking data. We also compare process variables in 465
 terms of their predictive effect as mediators. Finally, we conclude by conducting a small 466
 qualitative analysis of two dyads (one from each experimental group) to suggest mecha- 467
 nisms for explaining the main effects found. 468

Learning and collaboration 469

For the analyses related to our main hypotheses (learning gains, joint attention and quality of 470
 collaboration), we made sure that our Analyses of Variance (ANOVA) met the assumptions of 471
 normality and homogeneity of variance by generating and analyzing histograms and boxplots. 472
 We also made sure that our distributions did not have outliers beyond two standard deviations. 473

As predicted, we found that participants in the "visible-gaze" group outperformed the 474
 dyads in the "no-gaze" condition for the total learning gain: $F(1,40)=7.81, p<0.01$. For the 475
 sub-dimensions, they also scored higher on the transfer questions $F(1,40)=4.47, p<0.05$. The 476
 difference is likely to be significant for the terminology questions $F(1,40)=3.59, p=0.065$ 477
 and for the conceptual questions $F(1,40)=2.11, p=0.154$ with a larger sample, since the effect 478
 sizes are between medium and large (Cohen's d are 0.62 and 0.5, respectively). Additionally, 479
 we took students' GPA (Grade Point Average) into account to perform further analyses (four 480
 data points were missing, two in each condition). The difference between our two conditions 481
 remained significant when taking the GPA as a covariate: $F(1,36)=6.79, p=0.013$ and taking 482
 the dyads as the unit of analysis: $F(1,18)=9.19, p=0.007$. 483

The treatment group ("visible gaze") also had a higher quality of collaboration as measured 484
 by Meier et al. (2007) rating scheme. The total score is an average across the nine sub- 485

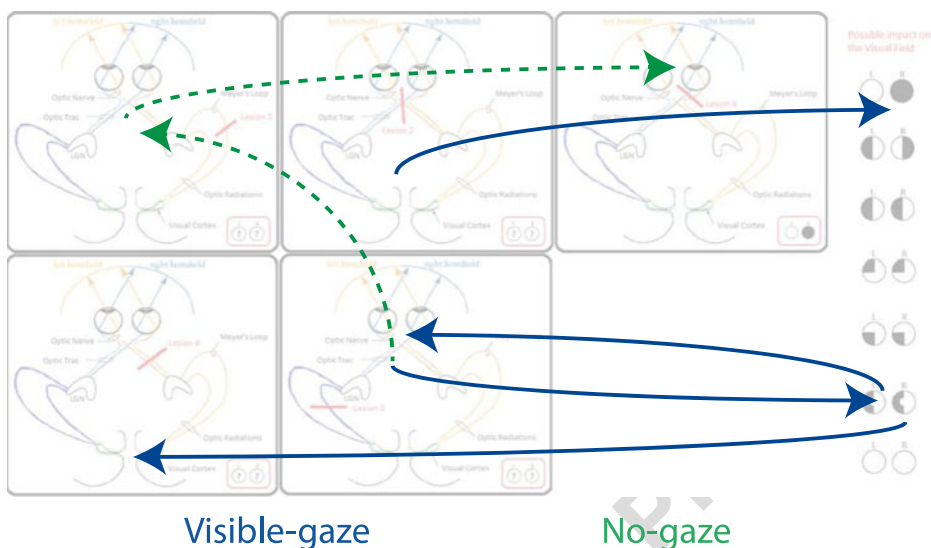


Fig. 3 Dashed arrows indicate that subjects in the “no-gaze” condition made more visual comparisons between two regions on the screen compared to subjects in the “visible-gaze” condition; *vice-versa* for the solid arrows

dimensions described in the “measure” section (as a reminder, each group was given a score between -3 and $+3$): $F(1,19)=11.73$, $p<0.01$, Cohen’s $d=1.24$ (mean for the treatment group= 0.89 , $SD=0.48$; mean for the control group= -0.08 , $SD=0.79$). More specifically, those visible-gaze condition dyads were better at *sustaining mutual understanding*: $F(1,19)=5.15$, $p<0.05$ (mean for the treatment group= 1.27 , $SD=0.88$; mean for the control group= 0.30 , $SD=1.03$), *pooling information*: $F(1,19)=7.53$, $p<0.05$ (mean for the treatment group= 1.18 , $SD=0.97$; mean for the control group= -0.20 , $SD=1.28$), *reaching consensus*: $F(1,19)=22.57$, $p<0.001$ (mean for the treatment group= 1.36 , $SD=0.79$; mean for the control group= -0.1 , $SD=0.55$), and *managing time*: $F(1,19)=4.98$, $p<0.05$ (mean for the treatment group= 1.00 ,

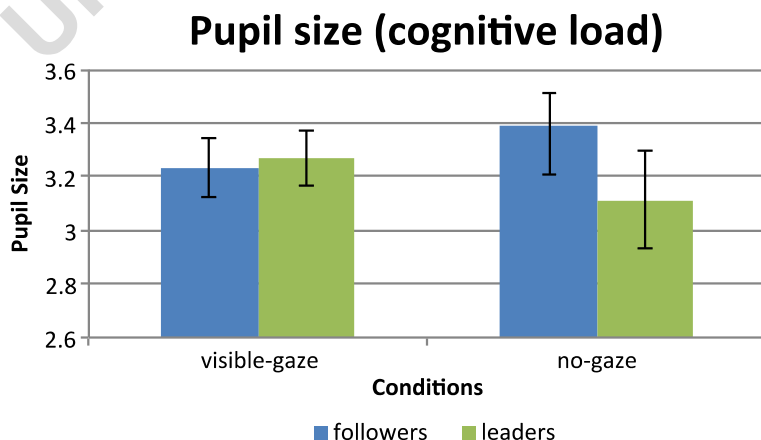


Fig. 4 Participants’ pupil size in each condition (distinguishing leaders and followers)

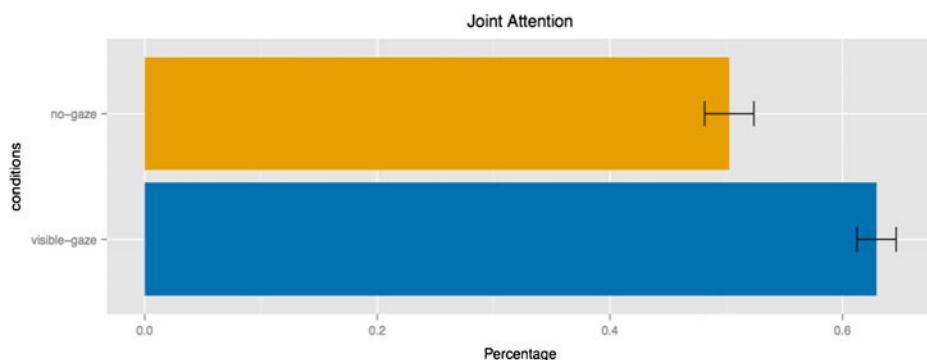


Fig. 5 Percentage of Joint Attention in each experimental condition

SD=0.67; mean for the control group=0.00, SD=1.29). A second judge double-coded 20 % of the video data; inter-reliability index using Krippendorff's alpha was 0.81. An alpha higher than 0.8 is considered as a reliable agreement between judges (Hayes & Krippendorff 2007).

We categorized each member of the dyad as 'leader' or 'follower' (Fig. 2). We found an interaction effect between those two factors (experimental condition and individuals' status) on the total learning score: $F(1,38)=5.29, p<0.05$. Followers who could see the gaze of the leader learned significantly more than followers who could not (see the "Total" column in Fig. 2). Overall, followers in the "visible-gaze" condition learned more than followers in the "no-gaze" condition: $F(1,19)=10.65, p<0.005$ (mean=0.54, SD=0.15, mean=0.35, SD=0.11, respectively). There was no significant difference between leaders in the two conditions: $F(1,19)=0.26, p=0.61$ (mean=0.44, SD=0.09, mean=0.42, SD=0.11, respectively). Additionally, leaders and followers did not differ in terms of their GPA: $F(1,36)<1$ (for followers, mean=3.19, SD=0.75; for leaders, mean=3.11, SD=0.76).

Eye-tracking Data We analyzed our eye-tracking data in order to describe the ways in which students' strategies differed when they could see the gaze of their partner on the screen while working on the contrasting cases. We excluded five subjects from those analyses because of missing data (due to the eye-tracker crashing during the activity). Three such participants were in the "no-gaze" condition, and two participants were in the "visible-gaze" condition. We thus have 37 subjects when measuring the number of fixations and saccades, and 16 dyads (32

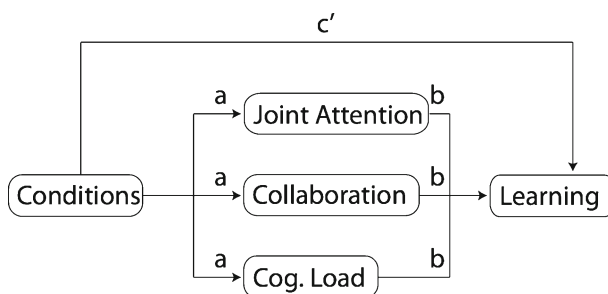


Fig. 6 Mediation model for our experiment. We tested the following potential mediators: cognitive load (measured by the size of participants' pupils), quality of collaboration (measured by Meier, Spada and Rummel's [2007] rating scheme) and percentage of joint attention (estimated with the eye-tracking data)

subjects) when measuring joint attention. We found that participants in the “no-gaze” condition had significantly more fixations on case one ($F(1,35)=9.69, p<0.01$, Cohen’s $d=1.28$), and case three ($F(1,35)=4.92, p<0.05$, Cohen’s $d=0.8$). Participants in the “visible-gaze” condition spent more time looking at the answers ($F(1,35)=10.41, p<0.01$ Cohen’s $d=1.21$).

In terms of examining the gaze saccades between regions, we divided the screen into six areas: the five contrasting cases and the answers on the right. Figure 3 summarizes the results obtained. Subjects in the “visible-gaze” condition made more comparisons between case five and the answers ($F(1, 34)=6.41, p<0.05$; Cohen’s $d=0.77$; $F(1,34)=7.14, p<0.05$; Cohen’s $d=0.82$), from the answer to case four ($F(1,34)=7.12, p<0.05$; Cohen’s $d=0.74$), and from case two to the answers ($F(1,34)=5.12, p<0.05$; Cohen’s $d=0.73$). Subjects in the “no-gaze” condition made more saccades from case one to case six ($F(1,34)=5.32, p<0.05$; Cohen’s $d=0.59$) and from case five to case one ($F(1,34)=6.14, p<0.05$; Cohen’s $d=0.81$). Even though we conducted multiple comparisons, we decided to follow Rothman’s advice (1990) to not adjust our results. This researcher suggested that not correcting results led to fewer errors and supported researchers’ explorations of alternative hypotheses. Additionally, since those results are descriptive and not central to our claims, we leave it to the reader to hedge their interpretation of those numbers.

We did not find a significant difference for cognitive load between the two gaze 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 effect between experimental condition and leader/follower status within the dyad is also not significant: $F(1,29)=2.51, p=0.12$, yet the effect size is between medium and large (partial eta squared=0.08; note that we are not using Cohen’s d , here), which suggests that followers in the “no-gaze” condition made more cognitive effort than followers in the “visible-gaze” condition. It would be interesting to have more subjects to see if this result becomes significant. The pattern is similar in direction to the one described for the learning test (i.e., followers tended to have a higher cognitive load than leaders in the “no-gaze” condition, and followers tended to have a lower cognitive load than leaders in the “visible-gaze” condition; Fig. 4).

Participants in the “visible-gaze” condition achieved joint attention more often than the participants in the “no-gaze” condition (see Fig. 5; for our analyses, we considered the percentage of moments of joint attention, in order to not give higher scores to subjects whose eyes were more easily detected by our eye-trackers): $F(1,30)=22.45, p<0.001$, Cohen’s $d=1.73$. This result holds when taking dyads (and not individuals) as the unit of analysis: $F(1, 14)=16.36, p<0.001$, Cohen’s $d=2.01$. The percentage of joint attention is one of the only measures correlated with a positive learning gain: $r=0.39, p<0.05$. Recall that our measure for joint attention was determined by whether, for each data point, the other member of the dyad was looking at the same area of the screen during the preceding or following two seconds, regardless of their verbal participation. We did not compute an exact match of x-y coordinates; instead, we considered each diagram and the column of answers to be areas of interest, since they have a conceptual significance for our task. Finally the percentage of moments of joint attention was also correlated with the quality of collaboration of the dyads: $r=0.58, p<0.001$.

Basic speech processing One explanation for students’ higher learning gain in the visible-gaze condition is that our intervention provided them with more opportunities for dialogue: By looking at their partners’ gaze, they could directly start a discussion on the diagram being looked at. Thus, to examine this conjecture, we analyzed the audio files of the experiment with a custom-made script that estimates the amount of speech produced by each participant. We found that subjects in the “visible-gaze” condition spoke more than the subjects in the “no-gaze” condition: $F(1,38)=6.13, p<0.05$ (mean=273.72 s., $SD=125.96$ for “visible-gaze”,

mean=189.11 s., SD=83.55 for “no-gaze”). This significant difference remains when taking the dyads as the unit of analysis: $F(1,19)=5.56$, $p=0.029$. At the individual level (when considering the amount of speech produced by each individual), this measure was not correlated with participants’ scores on the learning test: $r=0.24$, $p=0.14$. At the dyadic level (when considering the amount of speech produced by the group), this measure was associated with a higher percentage of joint attention: $r=0.46$, $p<0.01$.

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 cognitive effort exerted by the participants, or the percentage of joint attention for a dyad during the 12 min session may predict students’ learning. We tested for multiple mediation using Preacher and Hayes’ (2008) bootstrapping methodology for indirect effects. We used 5,000 bootstrap resamples to describe the confidence intervals of indirect effects in a manner making 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 ‘mediator’: Collaboration, percentage of joint attention, and cognitive load. GPA was used as a covariate, since our goal is to find mediators irrespective of participants’ grades. Results for multiple mediation indicated that only joint attention (CI: [0.03; 0.19]) was a mediator for learning (see Fig. 6).

Vignette As a reminder, we note that the previous sections provided quantitative data on the effect of a gaze-awareness tool on students’ remote collaboration. However, they do not provide us with any explanation for the mechanisms that may be responsible for such effects. Table 1 seeks to suggest qualitative explanations for the positive effect of our gaze-awareness tool on students’ learning gains and quality of collaboration. We compared two groups: one in the “visible-gaze” condition (Table 1 left side) and one in the “no-gaze” condition (Table 1 right side).

In terms of coordination, we found substantial differences between our two dyads. More specifically, the four following points summarize our qualitative observations:

- (1) First, 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 even mentioned the lesion’s number (v2). 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 gaze awareness intervention facilitated coordination and helped the follower anticipate the leader’s explanations.
- (2) Second, we found the emergence of interesting new anaphoric conventions: in the “visible-gaze” dyad (v3), when Lea says “so that would be... left-left, right-right”, neither of them explicitly stated that she was referring to the eyes and hemifields of the diagram. Rather, they implicitly built the convention of moving their gaze as a deictic gesture to complement their explanations—illustrating how conventions of efficient language use such as anaphora when individuals are co-located in a conversation (Clark 1996) extended to remote collaboration when an alternative referring mechanism (gaze in this case) can be used in the collaborative process of common ground construction during discourse.
- (3) Third, we hypothesized 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” (v2) suddenly made sense when Lea pointed her gaze at

Table 1

Visible-gaze (P54-P55)

No-gaze (P07-P08)

*Lea (L) is the leader, and Flo the follower (F).***(v1) 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.

(v2) 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.**(v3) 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 from the left hemifield to the left eye, followed by Flo's gaze]* and right-right *[gaze moving from the right hemifield to 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.

(v4) 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 eyeL: *[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]**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 between answer 4 and 5]*F: both the visions, they're blocked by one fourth. So it's not like completely blocked. So the answer would be that one *[gaze moving to answer 4]*.L: but how is it... hum... *[gaze still moving between answer 4 and 5]*. So you 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]*

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.

- (4) Fourth, we observed a tighter coupling between subjects' attention in the "visible-gaze" condition (v4): 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 talked, and then continue to explore other lesions (n3). This suggests that the theories built during the activity were more the result of the dyad's shared cognition (in the "visible-gaze" condition), and more the results of individuals' cumulative contributions in the "no-gaze" condition.

Discussion

Our findings demonstrate the importance of mediating technologies for 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. In both conditions, an audio channel was open between the participants in the dyad, as a medium for use in the coordination of their activities. Our results reveal that this simple intervention was associated with subjects in the visible gaze condition producing a higher quality of collaboration and learning more from the contrasting cases. In particular, participants characterized as 'followers' saw their learning gain dramatically increase as compared to 'leaders'. This result was partially confirmed by a similar pattern found for students' cognitive load: followers in the control group expended 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 one and six; 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. Interestingly, our measures of students' quality of collaboration did not significantly correlate with their learning gains. This suggests that perhaps visual coordination is a dimension of collaboration that is not captured by the rating scheme we used. Thus, our results may be an indication of future work for refining Meier, Spada and Rummel's approach for assessing collaboration.

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. Due to time constraints and limited access to participants, we did not include this condition in our experiment. However, we are interested in studying the effect of shared pointers in future studies compared to gaze-awareness tools. Conceptually, we believe that real-time mutual gaze perception may have 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 on the part of the user. There is also a certain degree of uncertainty associated with a cursor that stops

moving; is your partner thinking, being distracted, or waiting on 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.

There is an alternative interpretation for these results. Most participants were using an eye-tracker for the first time, and it is possible that this novelty effect generated more engagement toward the learning task in the “visible-gaze” condition. If this interpretation is warranted, it would to some extent undermine the usefulness of our results. However, by reviewing the transcripts, we found that only two dyads (17 % of the participants) commented on the gaze-awareness tool or expressed any kind of surprise. We plan on further exploring this question by using natural-language processing algorithms (NLP) on the transcripts and the audio data; more specifically, we will extract the arousal expressed by the participants’ voices (Boersma 2002) to see whether subjects in one condition showed more engagement toward the learning task. One should note, however, that differences in engagement can’t explain the differences found in the vignette above. Future work will explore how much variance of the learning effect each of those factors may explain.

General discussion

It is well established that joint attention plays a crucial role in any kind of social interaction. Our study provides additional evidence that its role is also preponderant in collaborative learning situations. We predict that in a near future, eye-trackers will become increasingly cheaper and widely available to a broad range of devices (e.g., not only desktops and laptops, but also smartphones and technology-enhanced eye glasses). Our study shows that in some technology-mediated interactions, real-time mutual gaze perception is beneficial for collaboration. Those results have important implications, especially for e-learning environments, since achieving a good remote collaboration is particularly challenging (Kreijns et al. 2003). Thus, we believe that it will be promising to explore the conditions under which students’ visual exploration should be made available to their partners when working remotely. One caveat is that this awareness tool seems to work well for dyads; we are more skeptical of the use of a gaze-awareness tool for triads or large groups where this visualization may become distracting. Future work should investigate whether this effect generalizes to different tasks and group sizes. Our findings also have indirect implications for co-located interactions; as Barron (2003) highlighted in her study, having students collaborate in the same space, either side-by-side or face-to-face, does not make the establishment of joint attention trivial. We hypothesize that our intervention may lead to similar benefits for students working on an interactive surface (as while wearing eye-tracking goggles). Finally, our results have further implications for teachers’ practices; with training, we posit that gaze-awareness tools could teach students the value of achieving joint attention in collaborative groups. The ability to effectively collaborate with peers was recently highlighted as a crucial 21st century competency (Trilling and Fadel 2009).

From the standpoint of developing the sciences of collaborative learning, this study has the merit of providing quantitative measures to advance Barron’s (2003) data-driven conceptualization of collaboration quality. In her study, she developed qualitative and quantitative indicators for joint attention from discourse and interaction analysis via intensive study of video recordings of collaborative learning dyads (for broader perspective on such analyses, see

Barron et al. 2013). In our approach, joint attention is computed in an objective way, without the need for a human coder. We argue that the use of eye-trackers is highly relevant for studying collaborative learning, because it provides the opportunity to speed up the research process by quickly computing metrics of interest and by providing real-time feedback to learners and teachers. With further developments, such measures could be used in real-time in classrooms; Liu et al. (2009) and Nuessli (2009) showed that machine learning techniques can predict students' expertise and task outcomes while the activity is still ongoing. In future classrooms, this approach could prove to be highly powerful in *sensing* and then *re-mediating* students' collaboration difficulties to prevent continued unproductive collaborative activities. Given real-time data processing (gaze and speech) and comprehensible data visualizations, teachers could immediately identify students who are having difficulties, and provide additional assistance to them. This kind of preventive approach is extremely appealing, because it allows teachers to provide help to students before they even realize that they need help. The employment of technological supports for overcoming collaboration challenges could also be triggered when such dyadic difficulties are sensed during eye-tracking, though at this stage of understanding, further studies are needed to guide the development of such real-time collaboration sensing and re-mediation scaffolding tools.

This study has several noteworthy limitations: (1) 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 where computer-using collaborators are side by side. One might assume that joint attention is more readily achieved in a face-to-face or side-by-side collaboration, but as we have noted, Barron's (2003) study with co-located middle school collaborators solving math problems indicated real collaboration challenges even when face to face and sharing documents for learning. Future studies using eye-tracking goggles on interactive surfaces could address this question. (2) Second, students had a limited amount of time to work on the contrasting cases. It was unclear how this limitation impacted students' performance. (3) Third, we only cursorily evaluated the transcripts of the dyads. More fine-grained coding schemes may provide additional clues as to how joint attention specifically facilitated collaborative learning; the interaction effect between followers and leaders is especially interesting, and should be analyzed in greater depth. (4) Four, our sample is rather small (especially when considering dyads as the unit of analysis); future studies should seek to replicate this effect with more subjects. (5) Fifth, 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. (6) Lastly, most of our statistics are conducted at both the individual and dyadic level. It would be more appropriate to conduct multi-level analyses to account for the learners' non-independence (Cress 2008). In particular, the fact that we analyzed eye-tracking data (fixations, saccades) at the individual level is a significant limitation of our work. However, since our claims are about joint attention, we made sure that the results related to this measure remained significant when taking dyads as the unit of analysis.

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 examine whether those processes (and others we may uncover) are significantly different across conditions. Second, a next logical step is to investigate this phenomenon in a different setting (e.g., in a co-located situation). Eye-tracking goggles could offer an interesting tool for this purpose. Third, it would be interesting to see if those results generalize beyond contrasting cases; it may

be, although we have our doubts, that this intervention is only effective for perceptual tasks. Finally, our results suggested that supporting joint attention between novices and experts could bring interesting results, as real-time mutual gaze perception provides a form of “inter-identity technology” (Lindgren and 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 and Brennan 2007).

Acknowledgments We gratefully acknowledge grant support from the National Science Foundation (NSF) for this work from the LIFE Center (NSF #0835854).

References

- Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE Transactions on Learning Technologies*, 5(3), 264–274.
- Aronson, E., Blaney, N., Sikes, J., Stephan, G., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills: Sage Publication.
- Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies*, 3(3), 203–213.
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55(4), 1278–1289.
- Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, 12(3), 307–359.
- Barron, B., & Roschelle, J. (2009). Shared cognition. In E. Anderman (Ed.), *Psychology of classroom learning: An encyclopedia* (pp. 819–823). Detroit: Macmillan Reference USA.
- Barron, B., Pea, R. D., & Engle, R. (2013). Advancing understanding of collaborative learning with data derived from video records. In C. Hmelo-Silver, A. O'Donnell, C. Chinn, & C. Chan (Eds.), *International handbook of collaborative learning* (pp. 203–219). New York: Taylor & Francis.
- Bates, E., Thal, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology*, 25(6), 1004–1019.
- Beatty, J. (1982). Task evoked pupillary responses, processing load and structure of processing resources. *Psychological Bulletin*, 91(2), 276–292.
- Beatty, J., & Lucero-Wagoner, B. (2000). Pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology*. New York: Cambridge University Press.
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Bransford, J., & Schwartz, D. (1999). Rethinking transfer: a simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education*, 24, 61–100.
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3), 1465–1477.
- Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, 35(1), 207–220.
- Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, 15(4), 481–498.
- 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. (pp. 173–180). New York, NY: ACM.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- 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.
- Cress, U. (2008). The need for considering multilevel analysis in CSCL research—An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning*, 3(1), 69–84.

- Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machine: Towards an interdisciplinary learning science* (pp. 189–211). Oxford: Elsevier.
- Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, 8(2), 145–157.
- 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(4), 596–615.
- Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3), 4.
- Jermann, P., Mullins, D., Nuessli, M.-A., & Dillenbourg, P. (2001). Collaborative gaze footprints: correlates of interaction quality. In Spada, H., Stahl, G., Miyake, N., & Law, N. (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings, Hong Kong, July 4–8, 2011, Volume I - Long Papers*, (pp. 184–191).
- Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.
- Kim, T., & Pentland, A. (2009, March). Understanding effects of feedback on group collaboration. In *Proc. of the AAAI Spring Symposium on Human Behavior Modeling* (pp. 1–6).
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19(3), 335–353.
- Leudar, I., Costall, A., & Francis, D. (2004). Theory of mind: A critical assessment. *Theory & Psychology*, 14(5), 571–578.
- Liu, Y., Hsueh, P.-Y., Lai, J., Sangin, M., Nussli, M.-A., Dillenbourg, P. (2009, June). Who is the expert? Analyzing gaze data to predict expertise level in collaborative applications. *Proc. of IEEE Int. Conference on Multimedia and Expo: ICME 2009* (898–901).
- 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(1), 63–86.
- Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, 20(1), 115–128.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–175.
- Pea, R. D. (1987). Socializing the knowledge transfer problem. *International Journal of Educational Research*, 11(6), 639–663.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 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(6), 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(3), 365–392.
- Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, 1(1), 43–46.
- Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational research*, 13(1), 89–100.
- Sangin, M. (2009). *Peer knowledge modeling in computer supported collaborative learning*. (Doctoral Dissertation). Retrieved from <http://infoscience.epfl.ch/record/133637>. Last access: 03/08/2013.
- Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences*, 4(3), 321–354.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition & Instruction*, 16(4), 475–522.
- Schwartz, D., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184.
- Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., & Wang, H. (2011, May). Enhancing genomic learning through tabletop interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2817–2826). New York: ACM.
- Sheingold, K., Hawkins, J., & Char, C. (1984). "I'm the thinkist, you're the typist": The interaction of technology and the social life of classrooms. *Journal of Social Issues*, 40(3), 49–61.
- Stem, D. (1977). *The first relationship*. Cambridge: 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: Erlbaum Associates, Inc.
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. New York: John Wiley & Sons.