Towards a Pedagogical Conversational Agent for Collaborative Learning: A Model Based on Gaze Recurrence and Information Overlap

Yugo Hayashi[1 2] **(y-hayashi@acm.org)**
[1]College of Comprehensive Psychology, Ritsumeikan University
2-150 Iwakura-cho, Ibaraki, Osaka, 567-8570, Japan
[2]Human-Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA,15213, U.S.A.

Abstract

This study focuses on collaborative learning involving a knowledge integration activity, whereby learner dyads explain each other's expert knowledge. It was hypothesized that learning gain can be determined by the degree to which learners synchronize their gaze (gaze recurrence) and use overlapping language (information overlap) during their interaction. Thirty-four learners participated in a laboratory-based eye-tracking experiment, wherein learners' gazes and oral dialogs were analyzed. Multiple regression analysis was conducted, wherein learning performance was regressed on the two independent variables. Then, a simulation was conducted to view how the model predicts performance based on the collaborative process. The results showed that both gaze recurrence and lexical overlap significantly predicted learning performance in the current task. Furthermore, the suggested model successfully predicted learning performance in the simulation. These results indicate that the two variables might be useful for developing detection modules that enable a better understanding of learner-learner collaborative learning.

Keywords: Collaborative Learning; Pedagogical Conversational Agent; Information Overlap; Gaze Recurrence

Introduction

Inspired by the social-constructive approach by Vygotsky (1980), numerous studies in the fields of cognitive science and learning science have investigated the mechanisms of how learners gain knowledge through social interactions with collaborative partners. In those studies, "knowledge integration" through constructive interactions, especially with partners who have different knowledge, provides opportunities for individuals to explain to others while reflectively considering their own perspectives (Aronson & Patnoe, 1997). Studies have pointed out that such learning helps in externalizing knowledge (Shirouzu, Miyake, & Masukawa, 2002; Miyake, 1986), facilitates meta-cognition during explanations (Chi, Leeuw, Chiu, & Lavancher, 1994) and perspective change (Hayashi, 2018).

Based on this prior literature, I focus on collaborative learning during a simple knowledge intergeneration task, wherein learners' task is to explain certain content to another learner with a different perspective on the topic. Particularly, I wanted to experimentally investigate the cognitive process of this interaction and develop a predictive model of learning performance. The ultimate goal is to apply such a model to develop automated learning support systems that can detect learners' activity and facilitate better peer collaborative learning.

Designing a Pedagogical Conversational Agent(PCA) for learner-learner collaboration

In the past several decades, numerous studies have been conducted with the aim of developing learning support systems in the field of intelligent tutoring systems (Koedinger, Anderson, Hadley, & Mark, 1997; Leelawong & Biswas, 2008). These investigations have illuminated how tutoring systems can be used to facilitate cognitive processes such as self-explanations (Aleven & Koedinger, 2002) and self-regulated learning (Graesser & McNamara, 2010). Some studies have examined how efficiently cognitive tutors can facilitate learning (Koedinger & Aleven, 2007) and investigations involving practical use of these systems have determine the types of knowledge that these systems best support (Koedinger, Booth, & Klahr, 2013).

Recently, an application called a Pedagogical Conversational Agent (PCA) has been introduced to tutoring systems, the result of rapid advances in artificial intelligence and network technology (Heidig & Clarebout, 2011). According to research on the development of such systems, learners can directly interact with virtual and autonomous tutors though a computer screen, providing learning prompts and metacognitive suggestions much like human tutors. PCAs have been demonstrated to be effective in the area of collaborative problem solving, such as for prompting achievement goals (Holmes, 2007) and providing periodic initiation opportunities (Kumar & Rose, 2011). A number of studies investigating the influence of PCA functional design have been conducted as well, focusing mainly on knowledge explanation tasks (Hayashi, 2012, 2014, 2016a, in press).

Despite the emerging body of literature on the design and development of PCAs for collaborative learning, a number of problems in this area remain unresolved, particularly the issue of how to detect learners' cognitive state and provide adequate feedback based on it. Investigations based on the methods used in cognitive science, such as modeling the cognitive states and interactions of collaborative learners, might be of use for designing efficient intelligent tutoring systems. Accordingly, the current study modeled the interactive process of learners during collaborative learning, with the aim of using the results to predict learners' cognitive states and incorporating this into PCA for automatic facilitation.

Capturing cognitive interaction using gaze recurrence and information overlap

Collaborative learning requires a communicative process such as establishment of a common ground (Clark & Wilkes-Gibbs, 1986), which plays an important role during explanation activities (Miyake, 1986). Taking this into account, I focus on two particular cognitive processes related to communication: learners' gaze behavior and language use.

Gaze Recurrence During conversations between speakers, it is important that the conversing parties refer to the same spatial referents (Schober, 1993) and use the same syntactic structures (Branigan, Pickering, Pearson, McLean, & Brown, 2011); it is also often important for them to physically synchronize in terms of gestures (Condon & Ogston, 1971) and simultaneously refer to the same referents (Richardson & Dale, 2005). Previous studies, such as (Richardson & Dale, 2005), have suggested that the degree of gaze recurrence between dyads (i.e., speaker-listener) is correlated with collaborative performance such as understanding and establishing common ground. They investigated how pairs engage in conversation through looking at a shared picture presented on a computer monitor. The researchers discovered that during these conversations, both speaker and listener physically gazed at the same area on the monitor. Moreover, using a technique called "cross recurrence analysis," they found that the gaze patterns of the dyad members synchronized over time. They also showed that the degree of synchronization correlates with comprehension tests such as memory retrieval and understanding. Studies on computer-supported collaborative learning (CSCL) involving gaze recurrence and techniques such as real-time mutual gaze perception have shown that gaze synchronization helps produce better collaborative learning (Schneider & Pea, 2014). The present study therefore focuses on gaze recurrence as one of the predictors of learners' performance because it aids learners in establishing common ground and making more efficient explanations during a collaborative task.

Information Overlap in Conversation Sociology research has conceived several basic principles for successful collaboration essentially, speakers should act cooperatively and mutually accept one another to be understood (Grice, 1975). In conversation, speakers implicitly adopt the same language as others in order to establish common ground, a process called lexical entrainment. This helps reduce ambiguity and ensure maximum clarity of reference between speaker and listener. The precise ways in which speakers agree on how a referent is conceptualized are called "conceptual packs" (Brennan & Clark, 1996). Taken together, these past studies have shown that to develop successful understanding through interaction, it is important for speakers to share the same knowledge. In fact, research has shown that speakers select common words when they believe that the others to whom they are speaking share their knowledge. Branigan et al. (2011) used a simple referential naming game, and found that speakers tend to try to align with each other even when interacting with a computer. However, another study investigating information overlap between speakers revealed that the more information people share, they less fluently they may communicate (Wu & Keysar, 2007). In any case, to successfully understand each other, speakers must exchange information on what they know and use that same knowledge to communicate. Therefore, in this study, when speakers refer to the same knowledge during their explanations, I considered them to be taking the other's perspective and trying to effectively coordinate with each other.

Goal and Hypothesis

The goal of this study was to understand the collaborative learning process of learner-learner dyads in order to automatically detect how successfully learners interact in a knowledge integration task by utilizing a PCA. To this end, I will focus on two indicators considered important for understanding communication in cognitive science: gaze recurrence and information overlap. First, I hypothesized that the learning process during a knowledge explanation task can be captured by the degree to which learners' attend to the same physical content (gaze recurrence) and their rate of overlapping knowledge (information overlap) during conversation. Second, I hypothesized that the efficiency of knowledge explanation (i.e., learning performance) can be predicted by a model of interaction that includes these two indices. To test these hypotheses, I conducted an experiment involving a learner-learner collaborative explanation task and analyzed the collected gaze and verbal data.

Method

Participants and conditions

Thirty-four (female: 15, male: 19, *M*age: 20.79, SD: 1.84) Japanese university students majoring in psychology participated in this study for course credit. This study was conducted only after ethical review and approval from the ethical review committee of the author's university.

Procedure

Upon participants' arrival to the experiment room, the experimenter thanked them for their participation and introduced them to their partner. The experimenter gave instructions on the task, which they were told was a scientific explanation task using technical concepts to explain human mental processes. Before the main task, they began with a free recall test about the concepts to check if they did not know any of the concepts that would be referred to in the task. Subsequently, they completed the main explanation task in about 10 minutes. After the main task, they completed another free recall test. Upon completing the entire experiment, they were debriefed.

Task

For the task, the dyad had to explain a topic in cognitive science (e.g., human information processing on language perception) using two technical concepts (e.g., “top-down processing” or “bottom-up processing”). As in the jigsaw method, which is often studied in the field of learning science and is a popular method of knowledge building in classrooms, I set up a situation wherein the learners did not know each other’s concepts. In other words, the experimenter provided only one of the concepts to each learner. Thus, to be able to sufficiently explain the topic using the two concepts, they would have to exchange knowledge by explaining the concepts to each other.

First, the learners had to explain each concept given to them to their partner. Information on this concept was provided to each learner before the task began. On starting the task, they were asked to first read the description and then explain what it meant to their partner. Learners were free to ask questions and discuss the concept with their partners. When one learner had finished explaining the concept, they switched roles and the other learner explained the given concept. The dyads were also instructed before the task that they would have to explain each other’s concept so that they could explain the topic using both technical concepts at the end of the task.

Experimental System

A redeveloped version of the system designed in a previous study was used(Hayashi, 2012, 2014, 2016a). Learners sat in front of a computer display and communicated with each other orally. The experimental system was developed in the Java language and worked on an in-house server-client network platform. The two learners’ computers were connected through a local area network, and task execution was controlled by a program on the server. Our version of the system also featured a PCA that provided meta-cognitive suggestions to facilitate their explanations.

During the task, the learners were not able to see each other and were instructed to look at the computer display while conversing with their partner. A brief explanation on the learner’s assigned concept was presented on each screen; their partner’s concept was covered so that they could not simply read and proceed individually. Accordingly, I expected that while one partner explained their concept by looking at the related screen area (Learner B), the listener (Learner A) would look at the same area. Two eye-trackers (Tobii X2-30) were used to collect gaze data. Furthermore, a microphone was placed next to each learner to record his/her voice through a two-channel mixer. All these audio data were transcribed manually.

An embodied PCA was presented in the center of the screen, which physically moved when it spoke. Below the PCA, there was a text box that showed messages. The experimenter sat aside in the experiment room and manually signaled the PCA to provide meta-cognitive suggestions. The timing of the suggestions was based on the following criteria: (1) whenever there was a gap in conversation, and (2) only once per minute. Five types of meta-cognitive suggestions are used, such as reminding learners to achieve the task goal (Azevedo & Cromley, 2004) and facilitating metacognition (Hayashi, 2014).

Masseurs

Performance The results of the pre- and post-task free recall test, wherein participants explained the topic using the two concepts, were used as the main performance data in this study. I coded the collected data according to how well the learners were able to explain their own concept as a result of successful coordination. The coding system was as follows: 0 = wrong, 1 = naive but correct explanation, 2 = concrete explanation based on presented materials, and 3 = concrete explanation based on the presented materials and using examples and metacognitive interpretations. While I also analyzed the partner’s explanation and the integrated explanation of the two concepts, these results will be mentioned in a later paper due to the length restrictions of this one.

(1) Gaze recurrence: Gaze recurrent analysis Gaze synchronization is important for successful communication. In the current task, the more the learners looked at the same area on their screen, the more attention they were considered to pay to each other (joint-attention). This can thus be interpreted as an index of perspective taking during the collaboration process.

Using the gaze data, I investigated the degree that each learner looked at the same area on the screen. The area was categorized according to the areas of interest, where the (1) area 1 = left frame box(self/other concept), (2) area 2 = right frame box(self/other concept), (3) area 3 = middle frame box(PCA), and (4) other. I labeled the fixation coordinates from each participant’s gaze log files from the eye tracking system corresponding to these areas.

Next, based on Richardson and Dale (2005), learners’ labelled data were analyzed using recurrence analysis, to capture the proportion of fixations at the same location for both learners in a typical time state. The analysis was conducted using R. The recurrence of *phi* observed between the two time-series (Learners A and B) is calculated for a specific time k. The $phi(k)$ coefficient increases with the frequency of matching recurrence at the same time (k; k) and decreases with the frequency of mismatching. Based on the procedure used by Richardson and Dale (2005), I adopted a time lag of 3 s. For each pair, I calculated the recurrence during each time lag, used the maximum value as the representative index for each learner and for further statistical analysis.

(2) Information overlap: Adopting an epistemic network analysis In the experimental task, learners must explain their own concept and try to understand their partner’s through conversation. Therefore, the more they use the same types of words during conversation, the more likely the learners are to be taking their partner’s perspectives to facilitate understanding. To analyze this, I used epistemic network analysis, a method of understanding the relations between coded data by representing them in a dynamic network. I have previously used this method to analyze the frequent use of important words during dyadic conversations (Hayashi, 2016b). The advantage of using epistemic network analysis in conversational analysis is that it can standardize the complexity of word types and enables comparison of word relations between speakers. For comparison, I calculated the similarity of word networks among learners and investigated to what extent they used the same words during the task.

(a) Development of a dictionary database The first stage of the analysis involved developing a dictionary database and collecting frequent words used during learners' conversations. I conducted a morphological analysis using the Rmecab package of R, extracting all nouns used more than twice. There were a total of 13,565 morphemes. Non-technical words (e.g., "I", "you") were deleted from the list manually.

Next, I developed a dictionary database of the extracted keywords for each dyad (Learner A and B). The average number of keywords for each pair was 34.70. An example dictionary was "processing," "perception," "knowledge," "experience," "instinct," "top-down," "bottom-up," "language," "friends," "relationship," "information," "explanation," "problem," "objective," "important," "occasionally," "elements," "current," "cognition," "input," "event." Subsequently, I generated a bipartite graph for each learner based on dictionary database of that pair, and calculated the degree of knowledge (i.e., word) overlap between the learners in the network.

(b) Network analysis For each learner, I developed a network based on the bipartite graph (i.e., n key words for X each learner's n utterances). Each node represents the lexical category of a keyword frequently used in each participant's explanation. The target of this analysis is to capture this degree of overlap and understand to what extent each dyad used the same knowledge during the task.

(c) Calculating the knowledge overlap In the next phase, I calculated the degree of knowledge overlap between the two networks. For this, I used the matching rate k, where l indicates the number of nodes used in both networks (Learner A and Learner B) and n stands for all the nodes in the network.

$$k = \frac{l}{n} \tag{1}$$

The closer the value of k is to 1, the greater the proportion of the same words used by learners during their conversation. This calculation was conducted for all pairs and used as an index of information overlap in the statistical analysis.

Results

Regression analysis of human data

Figure 2 shows the results of the Pearson's correlation analysis between each learner's performance (horizontal-axis) and the two main variables *phi* and k (vertical-axis). There was a significant correlation between performance and gaze recurrence *phi* (r=.385, p=.012) and a marginally significant correlation with information overlap k (r=.235, p=.090). Next, I conducted a multiple regression analysis where learning performance (dependent variable) was regressed on *phi* and k (independent variables). The regression coefficient R^2 was .439 and the F-value from an analysis of variance (ANOVA) was 3.709, indicating significance for both variables (p=.036). Thus, the two independent variables predicted learning performance(y). The regression equation is shown below:

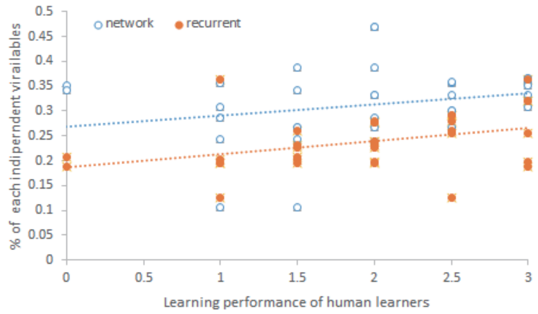

Figure 1: Relation between learning performance and each independent variable and *phi* and k.

$$y = -0.071 + (2.235 * phi_i) + (5.395 * k_i) \tag{2}$$

Using this model, I next wanted to see if it was possible to predict the learning performance of human learners via a simulation.

Predicting learning performance using suggested model

Next, using *phi* and k as parameters, I ran a simulation to see if the model can predict learning performance. This was conducted to see how accurately the system would detect the learning performance if the suggested model was implemented in the current experiment. Figure 2 shows the results of this simulation.

Discussion

To investigate the first hypothesis, learning process was captured through analysis of gaze recurrence and information overlap (*phi* and k). I found that learning performance was significantly predicted by using these two variables, thus supporting our first hypothesis. These findings suggest that the predictive regression model could function as a model of the learning process. However, the correlation between learning performance and k has rather weak (r=.235, p=.090). It therefore might be necessary to use a better lexical network analysis method, such as natural language processing techniques (e.g., correspondence analysis).

A weakness of this study relates to the collaborative learning setting, particularly, learners' lack of gaze awareness of

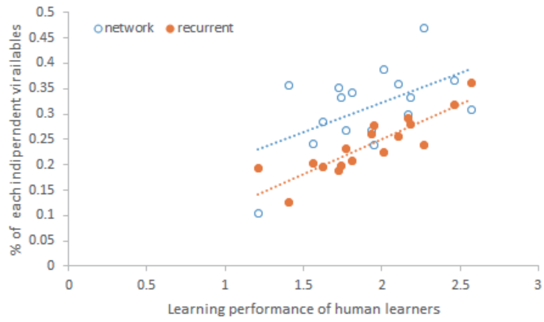

Figure 2: Predicting learning performance using the suggested model.

their learning partners during the task. As pointed out in (Tomasello, 1995), social partners must demonstrate awareness of what their partners are attending to, suggesting the need for an indicator of partners' gaze. Schneider and Pea (2014) investigated the influence of gaze on learning performance by using a colored dot to represent the partner's gaze. I have conducted a similar experiment, the initial results of which indicate that awareness of gaze might change interactive behaviors. The use of other sensor technologies could be applied to generate greater awareness of partners; however, this goes beyond the topic of this paper, so I would like to leave it for future studies.

The results of this study also reveal that gaze recurrence and information overlap can be used for developing a PCA that can automatically detect the learning process. The next step towards developing such a PCA would be dealing with some issues on real-time automation. Nevertheless, the PCA would require gaze analysis modules and voice recognition technology coupled with lexical network analysis, which is necessary to detect information overlap. Recently, technologies such as facial recognition have been used in intelligent tutoring systems to detect learners' emotional states (D'Mello, Olney, Williams, & Hays, 2012), and these could be used to detect the synchronized behavior of the learners. Combining these methods might enable the development of an efficient PCA for collaborative learning in the future.

Conclusions

Towards developing a PCA for collaborative learning, this study quantitatively captured the collaborative learning process of dyads. One difficulty in developing PCAs is to automatically detect to what extent learners are successfully interacting during a collaborative learning task. This study examined whether gaze recurrence and information overlap can capture the efficiency of coordination and thereby predict learning gains (in terms of their performance on an explanation task). The results showed that gaze recurrence and information overlap did indeed capture learning performance. Moreover, a simulation conducted using index data from actual learners as input information was able to reproduce the actual learner's performance. I discussed how these methods can be used for developing PCAs that can detect learners' cognitive interactive behavior and provide adequate facilitation prompts based on an evaluation of the system. This study might contribute to research on collaborative learning in cognitive science and has methodological implications on the design of PCAs for collaborative learning.

Acknowledgments

This work was supported by the Grant-in-Aid for Scientific Research (KAKENHI), No. 16K00219. I would like to thank Professor Kenneth R. Koedinger (Carnegie Mellon University) for his comments on this study.

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, *26*(2), 147–179.

Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom(2nd ed.): Building cooperation in the classroom.* New York: Addison Wesley Longman.

Azevedo, R., & Cromley, J. (2004). Does training on selfregulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, *96*(3), 523-535.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41 - 57. doi: https://doi.org/10.1016/j.cognition.2011.05.011

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482-1493.

Chi, M., Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Condon, W., & Ogston, W. (1971). The perception of language. In D. Horton & J. Jenkins J (Eds.), (p. 150-184). Columbus, OH: Charles E. Merrill.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: a gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, *70*(5), 377-398.

Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, *45*(4), 234-244.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3: Speech acts* (p. 41-58). New York: Academic Press.

Hayashi, Y. (2012). On pedagogical effects of learner-support agents in collaborative interaction. In *Proceeding of the*

11th international conference on intelligent tutoring systems(its2012) (p. 22-32).

Hayashi, Y. (2014). Togetherness: Multiple pedagogical conversational agents as companions in collaborative learning. In *Proceeding of the 12th international conference on intelligent tutoring systems(its2014)* (p. 114-123).

Hayashi, Y. (2016a). Coordinating knowledge integration with pedagogical agents:effects of agent gaze gestures and dyad synchronization. In *Proceeding of the 13th international conference on intelligent tutoring systems(its2016)* (p. 254-259).

Hayashi, Y. (2016b). Lexical network analysis on an online explanation task: Effects of affect and embodiment of a pedagogical agent. *IEICE Transactions on Information and Systems*, *E99.D*(6), 1455-1461.

Hayashi, Y. (2018). The power of a "maverick" in collaborative problem solving: An experimental investigation of individual perspective-taking within a group. *Cognitive Science*, *42*(S1), 69-104. doi: 10.1111/cogs.12587

Hayashi, Y. (in press). Multiple pedagogical conversational agents to support learner-learner collaborative learning: Effects of splitting suggestion types. *Cognitive Systems Research*. doi: 10.1016/j.cogsys.2018.04.005

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, *6*(1), 27-54.

Holmes, J. (2007). Designing agents to support learning by explaining. *Computers & Education*, *48*(4), 523-547.

Koedinger, K. R., & Aleven, V. (2007, Sep 01). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264. doi: 10.1007/s10648-007-9049-0

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education (IJAIED)*, *8*, 30-43.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935–937. doi: 10.1126/science.1238056

Kumar, R., & Rose, C. (2011). Architecture for building conversational architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, *4*(1), 21-34.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The betty's brain system. *International Journal of Artificial Intelligence in Education*, *18*(3), 181-208.

Miyake, N. (1986). Constructive interaction and the interactive process of understanding. *Cognitive Science*, *10*(2), 151-177.

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.

Schneider, B., & Pea, R. (2014). Toward collaboration sensing. *International Journal of Computer-Supported Collaborative learning*, *4*(9), 5-17.

Schober, F., M. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*(1), 1-24.

Shirouzu, H., Miyake, N., & Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science*, *26*(4), 469-501.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & J. Dunham P (Eds.), (p. 103-130). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Vygotsky, L. S. (1980). *The development of higher psychological processes*. Harverd University Press.

Wu, S., & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive Science*, *31*(1), 169-181.