Joint Action Theory and Pair Analytics: In-vivo Studies of Cognition and Social Interaction in Collaborative Visual Analytics

Richard Arias-Hernandez (ariasher@sfu.ca)
Simon Fraser University, SIAT, 250 - 13450 – 102 Avenue
Surrey, BC, V3T 0A3 CANADA

Linda T. Kaastra (lkaastra@magic.ubc.ca)
University of British Columbia, MAGIC, FSC 3640 - 2424 Main Mall,
Vancouver, BC, V6T 1Z4 CANADA

Brian Fisher (bfisher@sfu.ca)
Simon Fraser University, SIAT, 250 - 13450 – 102 Avenue
Surrey, BC, V3T 0A3 CANADA

Abstract

Herbert H. Clark's Joint Action Theory (JAT) has been groundbreaking for understanding the social and cognitive mechanisms that allow people to effectively coordinate joint actions in conversational, face-to-face settings. Using a method we call "Pair Analytics," we have extended the application of JAT to the study of analytical reasoning in computer-mediated, human-to-human interactions. Pair analytics (PA) sets a naturalistic scenario in which the social and cognitive role of human-human and human-computer interactions can be studied. In this paper, we support the claim that coupling JAT and PA is an effective research strategy to capture and study three socio-cognitive phenomena in collaborative visual analytics: (1) structuring and navigation of joint analysis; (2) management of joint attention; (3) and signaling of cognitively demanding tasks.

Keywords: Joint Action Theory, Pair Analytics, Visual Analytics, Analytic Reasoning.

Introduction

Joint Action Theory, Herbert H. Clark's theory of language in use (Clark, 1996), is a well-established psycholinguistic framework that has been very effective in bridging social and cognitive understandings of human communication. One of its basic tenets is that conversational, face-to-face, interaction is the foundation of human communication and language. Clark strongly criticizes theories of language that depart from this foundation by overemphasizing unilateral, cognitive, technological, or computational accounts that do not build upon the basic social structure of interaction that allowed language to emergence: face-to-face interaction. From a strong foundation on this theory of language, we have begun to expand the scope of Joint Action Theory from human-to-human interaction toward computer-mediated, human-to-human interaction. Our emphasis is on the use of Graphical User Interface (GUI) objects and Human-Computer Interactions (HCI) as gestures by which people mean things for others (Clark, 2005). The best way to illustrate this is with one example.

Imagine a scenario in which two data analysts share a computer screen and jointly work on the analysis of air traffic data. The layout of the screen contains several frames. Each one with a different visual representation of data for a commercial airline: a map, a table, and a bar graph. One of the analysts notices a salient bump in the bar graph and the following interaction unfolds:

> **Ben**: Look at this [Ben moves the mouse and places it on the bar in the chart that corresponds to "December"]
> **Anna**: well, it's December. You would expect a peak in air traffic.

In this interaction, Ben uses a deictic expression ("this") that requires attention to a specific referent on one of the visual representations of the data. In regular face-to-face conversation, deictic expressions are commonly used by speakers, along with non-verbal signals, such as finger or head pointing, to direct the attention of listeners to the referent implicit in the utterance (Clark, 2003). In our example, Ben uses the mouse as a pointing device to visually direct the attention of Anne to the exact place in the GUI where the object of the deictic expression can be identified without ambiguity. This behavior, equivalent to finger pointing, makes use of the mouse as a communicative on-screen gesture rather than as an input device of a human-computer interaction. In fact, the smaller size of the mouse pointer makes it more effective than using finger pointing to direct attention to a specific bar on a common 23-inch computer screen. The mouse-pointer is small enough not to block the general view of the bar graph or to point to more than one bar at a time. According to Clark's theory, this specific behavior uses an object of the GUI, the mouse, as a "material signal" (Clark, 2005), or as we refer to it in this paper, as a communicative "on-screen gesture." We claim that on-screen gestures are some of the most effective linguistic mechanisms that humans use to coordinate collocated, computer-mediated, human-to-human interactions. On-screen gestures, just as body gestures (Clark & Brennan, 1991), reduce the number of words and interactions that otherwise would be needed to communicate the same idea making communication less ambiguous and more effective.

This understanding of HCI in the context of human-to-human communication makes evident that interactive GUI components can potentially serve psycho-linguistic functions that have not been explored and are not currently well understood in the visual analytics domain. This paper presents an initial attempt to breach this gap by exploring the role of on-screen gesturing in three socio-cognitive phenomena of collaborative visual analytics: (1) structuring and navigation of joint analysis; (2) management of joint attention; (3) and signaling of cognitively demanding tasks.

Joint Action Theory and Pair Analytics

Joint Action Theory

Joint Action Theory (JAT) is a structured, socio-cognitive, theory of "language in use" developed by Herbert H. Clark (1996). For Clark, language use is an instantiation of a broader class of human practices: joint actions. In joint actions, individual participatory actions have to be coordinated to produce the intended effect. This implies coordinating *content* –what the participants intend to do, and coordinating *process* –how the participants effectively coordinate their individual actions to produce the desired joint effect. From this perspective, language in use is understood as a social process rather than as a mere exchange of information between speakers and listeners. For example, as a social process, the design of utterances is better approached as a participatory process that does not depend solely on the individual cognition and actions of the speaker but also on the cognition and gestures produced by the listener. When a speaker detects a facial gesture from the listener indicating confusion in the middle of an utterance, the speaker proceeds to elaborate or rephrase her wording without even waiting for an explicit, verbal request. Thus the construction of the utterance is better approached as a joint action between speaker and listener rather than the solely action of a speaker (Clark & Krych, 2004). The mechanisms that allow communication to be effective and reduce ambiguity and confusion are the result of a socio-cultural evolution of human language that has selected the most effective ones. Clark argues that humans constantly employ these mechanisms to solve coordination problems in joint actions (e.g. turn-taking, accounting for delays, navigating joint actions, mutual monitoring of understanding, sustaining joint attention, etc). Since collaborative visual analysis is an instance of a joint action, our starting theoretical point to apply joint action theory to visual analytics is that: *humans will use language and will work together to solve coordination problems in collaborative, visual analytics*.

Pair Analytics

In order to capture uses of language to solve coordination problems in collaborative, visual analytics, we have designed a method called "pair analytics" (Arias-Hernandez et al., 2011). Pair analytics (PA) is a method that generates verbal data about thought processes in a naturalistic human-to-human interaction with visual analytic tools. This method is loosely based on "pair programming" from "extreme programming" software development methods (Gallis, Arisholm & Dyba, 2003). Pair analytics requires a dyad of participants: one Subject Matter Expert (SME) and one Visual Analytics Expert (VAE). The dyad is given one analytical task, one data set, and one multi-screen computer with several visual analytics (VA) tools. The VAE has technical expertise in the operation of a suite of VA tools, but may lack the contextual knowledge that would be required to conduct meaningful analysis of the data set s/he is working on. The SME, on the other hand, has expertise in a specific analytic domain, but no knowledge of the VA tools. The pairing of SME and VAE is designed to generate a human-to-human dialog that makes explicit the mental models and cognitive processes of the SME and VAE during their collaborative visual analysis. For example, during the analytic interaction, the SME may provide expert knowledge to suggest visual comparison of relevant variables, detect patterns, and generate or test hypotheses. The interaction of the dyad with the VA tool also generates a human-artifact dialog in which machine-models interact with human mental models. For example, visualizations created by the dyad may result in unexpected outcomes that do not fit into existing mental models due to the way the VA tool handles the data. The analytical task and the dataset for pair analytics are selected from previous fieldwork studies of analytical work in the specific domain of expertise of the SME. Selecting a currently relevant analytical task and familiar datasets create a more naturalistic setting for observations of analytical reasoning. Interactions between participants, as well as between VAE and visual analytics tool are captured in video and screen capture.

There are several advantages that PA offers to cognitive science research in visual analytics with respect to other commonly used methods, such as protocol analysis (Ericsson & Simon, 1993) or ethnographic methods (Schneiderman & Plaisant, 2006). First, it is a non-intrusive method that takes advantage of the natural and continuous flow of speech necessary to coordinate joint actions (Clark, 1996). Since communication between participants flows continuously in PA there is no need for a researcher to prompt participants to keep talking, as they would do in think-aloud, protocol analyses. This addresses one of the limitations of "thinking-aloud" protocols in which once participants get immersed in the task, reduce or stop their verbalizations demanding the researcher to interrupt participants to resume their "thinking-aloud" (Trickett et al., 2000), affecting reasoning processes such as insight generation (Schooler et al.,1993). Thus, PA provides more complete data about analytical reasoning with less external intervention. Second, PA provides an empirical entry point to study not only individual cognitive processes but also social processes used to coordinate joint actions (e.g. use of gestures to signal delays in cognitive processes). In this aspect, PA gets closer to what could be achieved by ethnographic, studies of cognition "in the wild" (Hutchins, 1995). Similar to field studies, PA is conducted in-situ, in collaborative settings where domain experts normally conduct their analytical work. Thus, socio-cognitive behaviors that occur in collocated, collaborative work settings also occur in pair analytics. An advantage with respect to ethnographic field studies of cognition is that PA maximizes the richness of the data being captured for content analysis by using screen and video capture.

Using JAT to analyze PA data

We use the video and screen data collected in the PA sessions to transcribe and code all conversations, joint actions and HCIs. First, we focus on transcribing all of the speech, verbal and non-verbal gestures used by participants of pair analytics. Second, using screen capture data, we complement the initial transcription with all of the human-computer interactions. Finally, we separate the transcription as sequences of joint actions, the basic analytical unit in JAT. Clark's methodology requires human-to-human conversation to be structured as a succession of hierarchical joint actions; each one with an entry, a body, and an exit (Clark, 1996; Bangerter & Clark, 2003). After organizing the transcripts in the hierarchical structure of joint actions, we move to a coding phase.

From an extensive review of the literature on JAT, we drafted an initial coding scheme to capture three socio-cognitive phenomena in collaborative visual analytics: navigation of joint actions between different analytical phases, coordination of joint attention, and use of gestures to signal delays in joint actions produced by cognitive workload (Arias-Hernandez et al., 2011). Using the coding scheme, we code several pair analytic sessions and analyze the results.

Pilot Study

To test and refine this theoretical and methodological approach for the study of psycho-linguistic mechanisms used by analysts to solve coordination problems in collaborative, visual analytics, we conducted a pilot study.

Setting:

Our study involved four subject matter experts (SMEs) in aircraft maintenance engineering and two visual analytics experts (VAEs) from our laboratory. In collaboration with the SMEs, we agreed to work on a real analytical task that the aircraft maintenance analysts were struggling with at the time. The analytical task was open-ended and loosely structured. The objective was to generate and test hypotheses that could explain differences of unscheduled aircraft downtime by models of aircrafts in a commercial fleet. The maintenance dataset to be used for the analysis was structured and contained 45 fields and over 90,000 records. One pair analytics session was arranged for each SME, for a total of four sessions. Pair analytic sessions were conducted in-situ, over a period of four weeks in September and October, 2009, and sessions had an average duration of 2 hours. Tableau®, a visual analytics tool, was chosen by the visual analytics experts to be used in the pair analytic sessions. Since the visualizations generated by Tableau® during the sessions are mostly line and bar charts, no especial training was required for the SMEs to understand the visual representations of data. However a general introduction about the structure of the data and Tableau® was provided to each subject matter expert at the beginning of each session.

Results

Our results showed that in effect, communication between participants flowed continuously during the pair analytics sessions, and there was no need for a researcher to prompt participants to keep talking about their strategies, methods, or findings. We observed no decreasing in the amount of verbalizations, even when participants were engaged in cognitively demanding tasks.

The analysis of the JAT-informed coding scheme was organized around three socio-cognitive phenomena:

Ad-hoc structuring and navigation of the pair analysis: One classic study by Bangerter & Clark (2003) demonstrated that in American English people structure and navigate joint activities by using vertical and horizontal markers. "Vertical markers" are verbal gestures, such as "okay," and "all right," that signal transitions *between* joint activities. "Horizontal markers," such as "yeah" and "mhmm," on the other hand, are used to signal continuation *within* a singular joint activity (Bangerter & Clark, 2003). By using vertical and horizontal markers, people create ad-hoc structures of joint actions and navigate through them in an orderly fashion. In our analysis of the pair analytics sessions, we found ample evidence of the use of vertical and horizontal markers to navigate pair analytics (Arias-Hernandez et al., 2011). Moreover, we also found that the resulting structure being produced by the use of markers clearly distinguished the different analytical strategies, methods, and findings of the analysis. Using these markers, we were able to map all of the different lines of reasoning being pursued by the participants. We called these lines of reasoning that corresponded to distinctive joint activities: "analytical paths." Each analytical path corresponded to a complete form of joint action, with markers to signal its entry, its body, and its exit. The structure of analytical paths corresponded to a tree-like structure.

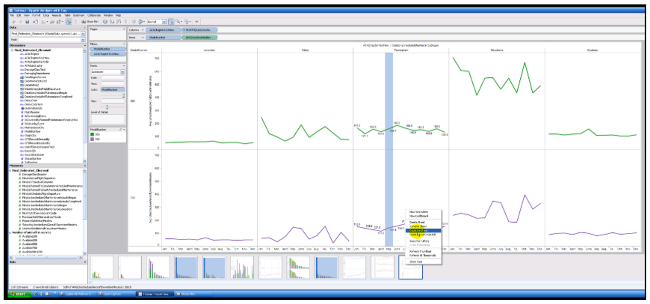

Figure 1. Using "placing-for" gestures in Tableau®.

During the analysis of human-computer interactions, we also found that participants accompanied verbal, vertical markers with non-verbal, on-screen gestures. For example, Tableau® provides a history feature that allows users to create a visual timeline of different views that have been generated during the course of the analysis. Every time users want to save a view for later recall, they create a snapshot of the current screen by either generating a new worksheet or duplicating the currently active worksheet. This action saves the current state of the analysis and generates a new thumbnail representing the next analytical project (Fig. 1). This interaction with the tool is functional, since its purpose is to save the state of the analysis and create a linear, visual history of several states saved. However, during our analysis we found that the VAE consistently produced this behavior in response to uses of vertical markers, such as "so," and "okay," visually signaling a vertical transition between different analytical paths. In doing so, her interaction also had a communicative effect. By producing the interaction on a visually shared space and by timing it with vertical transitions, the VAE communicated to the SME her tacit understanding that a vertical transition was in effect, and that she was ready to move to a next phase of analysis. We conceptualized this "placing-a-new-thumbnail-in-the-history" behavior as an "interactive marker," or a computer-mediated marker. Interactive markers are a subclass of on-screen gestures, and non-verbal gestures, whose purpose is similar to that of vertical and horizontal markers. According to Clark (2003, 2005), this particular interactional marker corresponds to the sub-category of "placing-for" gestures, since it visually "places" a new snapshot in the history line. Its use demonstrated that participants of pair analytics extended their repertoire of body-centric vertical markers to incorporate human-computer interactions and GUI elements as "material signals" (Clark, 2005) to help them navigate their joint analysis. Whether an insight was generated in the conclusion of an analytical path, or a dead point was reached, this on-screen gesture, "placing-a-new-thumbnail-in-the-history," served to communicate the tacit understanding that a milestone in the analysis had been reached and that a new analytical path was about to begin.

Management of joint attention: Coordinated joint attention is a pre-requisite for successful participation in a joint activity. Since participants in joint activities continuously propose joint projects to each other, the attention of each participant needs to monitor the continuous flow of signals. If attention is not focused on the relevant signal, then the intention behind the signal will not be communicated and the joint action will fail (Clark, 1996). In face-to-face settings, participants establish that joint attention is in place through the use of salient perceptual phenomena, perceptual co-presence, and gestural indications such as tone of voice, mutual gaze, finger pointing, and verbal markers (Clark, 2003). One of the results of our analysis of on-screen gesturing was that participants used "mouse pointing" as an extension of more traditional face-to-face mechanisms, either to direct the attention of the listener or to confirm to the speaker that attention was in the right place. In both cases, "mouse pointing" was used as a "directing-to"(Clark, 2003) kind of signal. While in both cases the mouse was in the hands of the VAE, it was used for different purposes. In the first case, the VAE, acting as the speaker, used the pointer to direct the attention of the listener (SME) to a visual object on the screen. On the other hand, in the second case, the VAE, acting as the listener, used the pointer to direct the attention of the speaker (SME) to the visual object on the screen where the attention of the VAE had been directed to by SME's speech.

The first case is equivalent to finger pointing or head pointing in face-to-face interaction. Its purpose is to indicate the location of a referent mentioned in speech. In all of its instances it was executed by the VAE as speaker. Attention to this signal and proper identification of the object being signaled was necessary to eliminate ambiguities in the use of demonstrative pronouns (e.g. this, that) and adverbs (e.g. here, there) when referring to visible objects in the GUI. In all of these gestures, the mouse was used to communicate rather that trigger events in the visual analytics tool. In other words, "mouse pointing" corresponded to a computer-mediated human-to-human interaction rather than a human-computer interaction. Thus, it was better understood as an on-screen gesture.

The following is an excerpt taken from the transcript of one of the sessions that illustrates this first kind of "directing-to" use of mouse pointing:

SME: okay
VAE: [clicks on the orange section of the bar "HOU"]
lots leaving from Houston
SME: So, that's interesting
VAE: yeah ... [clicks on the orange section of "DAL"]
lots leaving from Dallas [clicks on label of the bar
"DAL" on the X-axis]
SME: yeah

The second case of mouse pointing is observed when the VAE is the listener, not the speaker. Its purpose is to provide confirmation that the listener's attention is directed to the location or object where the speaker intends it to be. In face-to-face interaction, gaze and body position fulfill a similar function. When the speaker points to an object in her speech, she expects the listener to orient her body and gaze towards the object. This provides visual confirmation that joint attention is in place (Monk & Gale, 2002). However, in our pair analytics data, we found that mouse pointing provided a more nuanced and a more precise visual confirmation than that provided by gaze or orientation. When the speaker (SME) was referring to an object on the screen (sometimes pointing at it, sometimes not) the listener (VAE) would use the mouse to point at the object being discussed. This "confirming" behavior produces a visual cue that informs *exactly* where the listener's attention is located on the visual display. Gaze and orientation provide a more general, but less precise, visual cue. Due to the many visual features, objects, and the size of the screen, gaze and orientation do not afford the same precision as mouse pointing, which confirms *exactly* the object to where attention is directed. The following excerpt illustrates an instance of "confirming" joint attention with the mouse:

SME: so ... looking at that ... let's see the ... 200s are the orange
VAE: [moves the mouse over one of the bars with a visible orange stack] [inclines his head to read the vertical labels] yeah, so ...

"Self-talk" and on-screen gestures inform about cognitively demanding tasks: During our data analysis, we noticed that some of the joint activities were temporarily paused by one of the participants. The pauses, however, were characterized not by participant reducing their verbalizations, but rather by participants switching to "self-talk." Once the pause was finished, the participant would resume her participation in the joint activity. For example, occasionally the VAE would get a request from the SME to create a non-trivial view of the data. In response, the VAE would interact with the computer in solo mode, using self-talk, and conducting several steps to produce the intended view. Once done, the VAE would resume conversation with the SME. The SME, on the other hand, also engaged in similar kinds of behavior. When observing a new view of the data, the SME would stop interacting with the VAE to observe features of the view, use self-talk, and return to conversation afterwards. The following excerpt from one of the sessions illustrates the VAE using self-talk in one pause that lasted almost 11 seconds (self-talk in bold):

VAE: **why?** [unchecks "300" from the checkbox of filters, leaving only "700" checked. The colors change to purple from green but the view remains unaltered. Checks and unchecks "300" twice more] (9 sec) … **why those** [using his palm to cover his face] **overlap like that?** (2 sec)

It is important to note that these pauses are not interruptions in the activity since both participants are still on-task and advancing the joint activity. These pauses are better conceptualized as delays caused by the cognitive demands on participants generated by the ongoing task (Smith & Clark, 1993). As Smith and Clark (1993) have noted, the social substratum of joint actions demands that participants inform each other about problems that they encounter in their interaction, such as cognitively demanding tasks (e.g. retrieving information from memory, understanding difficult questions, etc). When faced with these cognitively demanding challenges, participants delay their response to the other participant's original request. However, due to the social context in which interaction occurs, the timing of response is crucial. Any delay in responding is open to several interpretations by the requester, some of which could undermine the responder's self-presentation (Goffman, 1978). In order to "save face" (Goffman, 1978), people normally resort to the use of fillers, such as "uh" and "um" to account for shorter delays, and self-talk to account for longer delays (Smith & Clark, 1993). According to Smith and Clark (1993), self-talk is a strategy used in conversational settings to (1) inform about delays and (2) inform about engagement in the joint action. Based on this rationale, we coded for "self-talk" moments in our data and mapped the activities that participants were doing during these moments (e.g. task and duration of the task). We found that most of these activities participants were engaged with during self-talk corresponded to human-computer interactions. So, we decided to categorize, time and analyze each of these by participant.

We found that different than Smith and Clark's studies on answering questions, human-computer interactions in pair analytics during self-talk have the additional advantage of a shared visual space (i.e. the interface) that provides additional information about the progression of the task while the delay is still in place. Every human-computer interaction that co-occurred with self-talk, visually informed the requester about the progression of a cognitively demanding task that was being executed by the responder. In other words, the combination of self-talk and HCI, not only served the two purposes theorized by Smith and Clark, but also served another function: (3) *to visually inform about the progression of the cognitively-demanding task that originated the delay*.

Our analysis also showed that *all* of the VAE activities during self-talk involved HCIs. For example, in one instance of an activity that we coded as: "confronting anomalies in a generated view," the VAE had created a bar graph with information about air traffic in origin and destination airports. On the x-axis of the bar graph, the VAE plotted data by origin and destination airports (two variables), and on the y-axis, she provided a count of annual flight from one specific origin to one specific destination (one variable). The SME asked the VAE to sort the bar graph by origin city (request), and the VAE proceed to select the view and clicked on the "sort descending" button (initiates response). However, the resulting view did not corresponded to the expected sorted result triggering the VAE to initiate self-talk (delay) while figuring out what was going on:

VAE: **I don't know how it is sorting it there** [clicks on the Y-Axis, selects the whole graph, clicks on sort-descending icon, updates the view, gets the same result] (7-sec delay)

Our analysis showed that this unexpected result occurred because two variables had been plotted on the x-axis creating sub-groups of bar graphs: first, by destination city, then, by origin city. In effect, sorting was occurring, but its expected visual result (gradually decreasing bars) was not visibly salient. In this case, the tool was sorting by the total values of the first variable (destination city), in a visualization that had this variable disaggregated by a second variable (origin city). In other words, sorting was occurring at the aggregated level of destination city (not visible), and not at the more disaggregated level of origin city (visible). One of the visual advantages of sorting simple bar graphs is that visual perception quickly processes the differences in size between bars, reducing the cognitive load of trying to determine the same differences in a non-sorted bar graph. However, in this case, sorting views with more than one variable per axis in bar graphs did not produce "visually-salient" and "intuitively-sorted" bar graphs, creating confusion and generating a delay. Here, the unexpected visual outcome of the sorting process and the consequential interactions to "fix it" were adding more cognitive load rather than reducing it. This example illustrates how analyzing the human-computer interactions that co-occur with self-talk can point out to concrete instances in which the visual analytic tool is not reducing the cognitive demands of the task at hand, but rather creating or increasing these demands.

Conclusion

Joint Action Theory (JAT), the Pair Analytics method (PA), and the results of the pilot study presented in this paper show that the JAT/PA research strategy provides a novel and original approach to understanding some of the psycho-linguistics mechanisms that analysts deploy to solve coordination problems in collocated, collaborative visual analytics. Future research using the JAT/PA research strategy will address more directly the specific kinds of affordances that visual analytics tools offer to enable users to navigate analytical paths and mark milestones in their analysis (e.g. structured history of the analysis process marking precise moments of insights). We will also continue to explore the role of "self-talk" events during pair analytics as indicators of cognitive demanding tasks and to differentiate between cognitive load caused by the demands of the analytic task from those caused by poor interface design. Experimental studies will also be conducted to test the efficacy of pair analytics in comparison to protocol analysis for keeping constant the flow of verbalization, even during cognitively intensive tasks.

Acknowledgments

This research was supported by NSERC STPGP 336506 "Visual Analytics for Safety and Security," MITACS ACCELLERATE BC "Program to Evaluate and Improve Visual Analytic Processes for Analyzing Maintainability, Reliability and Safety Data," and "Boeing Support for Visual Analytics in Canada" (Fisher subgrant).

References

Arias-Hernandez, R, Kaastra, L.T., Green, T.M. & Fisher, B. (2011). "Capturing Reasoning Processes in Collaborative Visual Analytics," *Proc. HICSS-44* (CD-ROM), Computer Society Press (2011), 10 pages.

Bangerter, A. & Clark, H.H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27 (2), 195-225.

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H.H. (2003). Pointing and placing. In: Kita, S. (ed.) *Pointing: Where language, culture, and cognition meet*, Psychology Press, 243-268.

Clark, H.H. (2005). Coordinating with each other in a material world. *Discourse studies*, 7 (4-5), 507-525.

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

Clark, H.H. and Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50 (1), 62-81.

Ericsson, K.A. and Simon, H.A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge: MIT Press.

Gallis, H., Arisholm, E. and Dyba, T. (2003). An initial framework for research on pair programming. *Proc. ESEM*, IEEE Computer Society.

Goffman, E. (1978). Response cries. *Language*, 54 (4), 787-815.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.

Monk, A.F. and Gale, C. (2002). A look is worth a thousand words: Full gaze awareness in video-mediated conversation. Discourse Processes, 33 (3). 257-278.

Schooler, J.W., Ohlsson, S. and Brooks, K. (1993). Thoughts beyond words: When language overshadows insight, *Journal of Experimental Psychology*, 122 (2), 166-183.

Shneiderman, B. and Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. *Proc. AVI 2006*, ACM Press, 7-12.

Smith, V.L. and Clark, H.H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32, 25-38.

Trickett, S.B., W.T. Fu, C.D. Schunn and J.G. Trafton, (2000). From dipsy-doodles to streaming motions: Changes in representation in the analysis of visual scientific data. *Proc. CogSci 2000*. Erlbaum.