INTERACTIVE TEAM COGNITION: DO GAZE DATA ALSO TELL THE STORY?

Dirk Schulze Kissing
German Aerospace Center (DLR)
Hamburg, Germany
Carmen Bruder
German Aerospace Center (DLR)
Hamburg, Germany

Gaze-data could be feasible to assess interactions within small groups, and provide added value for the assessment of 'team cognition' (c.f., Cooke et al., 2013). In the synthetic Control Center Task Environment (ConCenT) 18 teams of three are collectively monitoring an array of displays and predicting malfunctions, indicated by a setpoint kickbacks. To locate potential malfunctions they have to collaboratively determine system dynamics patterns. Within-team expectations are measured by gaze parameters related to situational relevant areas of interest (AoI). Within-team standard errors (SE) in fixation frequencies are utilized for post-hoc classifications of teams according to gaze-behavior homogenity. The amount of attentional resources teams allocated to relevant AoIs increases during critical phases, with a change of emphasis between the three relevant task elements. Post hoc groups do not differ in their way of monitoring relevant elements. It is concluded that gaze-data provide promising measures of interaction-patterns for team cognition analysis.

The concept of "team cognition" (see Cooke, Gorman, Myers, & Duran, 2013) provides a system-theoretical approach to the analysis of cognitive requirements in collaborative working environments, like airline operational control center (OCC) or area control center (ACC) of air navigation service providers (ANSPs). While the methodological approach to the study of shared mental models (e.g., Cannon-Bowers, Salas, & Converse, 1993) lies in the reproduction of the individual knowledge representations of the team members, whose consistency is subsequently examined, the research on the team cognition starts with the recording of the interactions developing over task performance. The basic theoretical assumption is that teams can be viewed as systems in which the phenomenon of team cognition emerges. This phenomenon does not take place at the level of the representations of individual agents, but rather materializes in the interaction patterns of individuals. The individual behavior, in turn, is affected by these emergent behavioral patterns at the team level.

The phenomenon of team cognition, as Cooke et al (2013) continues, unfolds over time, with interaction patterns changing within shorter periods of time than individual behavior patterns do, which are supposed to be more strongly determined by individual knowledge structures. Cooke et al (2013) therefore assume that individual behavior patterns change more slowly than team behavior patterns.

Cooke et al. (2013) postulate that team cognition should be measured in the context of unfolding task processing, with a focus on the processes running at the team level rather than on relatively static representations. Eye-tracking technology can provide such within-task objective behavioral measures for determining the quality of team performance (see e.g., Hauland, 2008). Cooke and Gorman (2009) also discuss eye movement measurement as a method for recording event data that can be used to systematically describe interactions in teams.

The main goal the presented work is dedicated to is the recording of team cognition as an emergent dynamic activity (Cooke et al., 2013) using integrated visual data measurement. We are looking for a metric to visualize coordination patterns within groups and to separate them from individual behavior patterns. For this purpose a synthetic task environment is developed.

The Synthetic Control-Center Task Environment ConCenT

According to Hess et al. (2005), the main features of a synthetic task environment should be analyzed both in the field of task-work as well as from a cognitive perspective (see Cooke & Shope, 2004). Therefore the design of the *Synthetic Control Center Task Environment* (*ConCenT*) was well-based on field observations of task-work in operational control centers, and by the outcomes of subsequent counseling workshops with operational experts from various fields, but in particular from the aviation domain (cf. Schulze Kissing & Eissfeldt, 2015 for further information).

ConCenT (Schulze Kissing & Eissfeldt, 2015; Schulze Kissing & Bruder, 2016) was designed to assess coordinative behavior within small groups (of N= 3 members) collectively working on management by exception scenarios (cf., Dekker & Woods 1999). The main task is to monitor an array of displays and to detect a malfunction, indicated by a setpoint kickback, intime (i.e., within 4 seconds after kickback-occurrence). Before the critical phase of potential malfunction the location of its potential occurrence can be predicted by the team if it exchanges and interprets relevant information on certain system dynamics. During the upcoming critical phase the individuals' reliance on where to expect a malfunction-event is then measured by gaze parameters related to situative relevant areas of interest (AoI). *ConCenT* emulates an operations-control center (OCC) at a company headquarter where at three working positions the output of three production plants (denoted by numbers one to three) at different company sites (denoted as Alfa, Bravo and Charlie) are remotely supervised. At the fields, direct control of the plants, plant supervisory, production control and production scheduling is assumed to be under fully automatic control. The task of the human operators in the OCC is to manage by exception. Each company site features the same layout (compare Figure 1): one power-station is providing energy for the three plants. Each plant represents a production system consisting of seven units, three assembly lines and four production units providing the components they need. In the three assembly lines different combinations of three out of the four components are assembled to dissimilar end-products. In total the the team has to supervise the outputs of three assembly lines at each of the three plants at each of the three sites (i.e., 3x3x3=27 production-process outputs). To enhance coordination demands, and thus promote interaction within the OCC team, responsibilities are subdivided. A human operator of the OCC is in charge to monitor that output-quantities match the presets for one out of the three assembly lines at each of the plants at each site (so each human operator has to supervise 1x3x3=9 production-process outputs; compare Figure 1).

Working positions

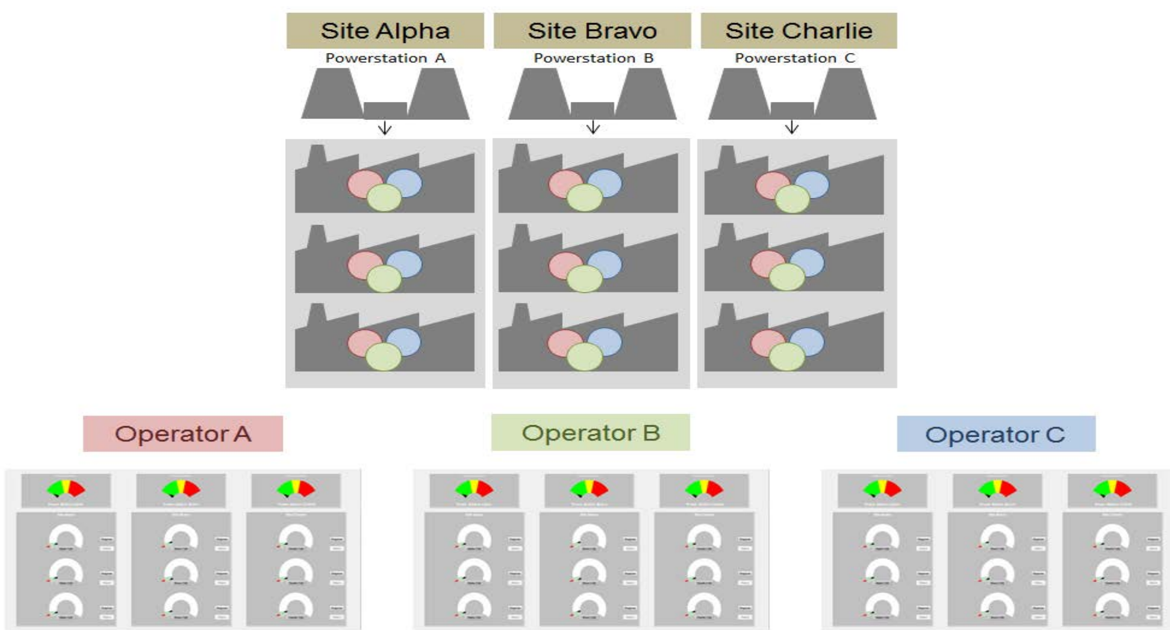

Figure 1: *Top:* Logical structure of ConCenT; three sites with three plants containing a production system with overlapping assembly lines controlled by different human operators; *Below:* User Inferfaces; each participant monitors 12 gauges, nine production-output and three power-supply displays, in a spatially corresponding array; at the output-gauges red arrows designate the actual setpoint, black arrows the actual production output, and green fields the range of tolerance for mismatches. At the power gauges black arrows display the current energy demand, and the coloured fields designate criticality.

Figure 1 presents the workstation for an operator with 9 gauges displaying the actual production output. The gauges are arranged in a 4x3 matrix correspondingly, with each column displaying the incoming information from only one site (Alfa, Bravo or Charlie), with the upper line showing the gauges displaying the actual energy consumption at the site, and each line below displaying the incoming information from the plants according to their numbers (1 to 3). Next to the right of each gauge a related response button labeled “Diagnose” is displayed.
The information stemming from the power-plants is identical at all working positions. The information stemming from the plants differ between working positions, because each gauge is displaying the output of another assembly line of the production system that is set up at that plant. So each operator has a different window to the production systems the OCC supervises. The operators therefore need to exchange their distinct information to get the whole picture.

Scenario

Sequence of phases. The whole system shows repetitive dynamics, shaped by assumed automatic control loops recurring every 30 seconds. During this cycle participants observe the events unfolding in the following sequence of phases: 1) *Notification* about a production schedule event: beginning with changing set-points at one or more gauges; this can be considered to be a notification to each team member where in his or her area of control a new set point is triggered and thus automation control is about to set in (duration: 6 seconds); 2) *Adjustment*: The remote automation adjusts production levels to the new presets; this can be observed only by the human operator in charge; energy is expended to perform this control process; thus the automated adjustment is accompanied by a temporary rise in energy demand. (duration from start of change: 9 seconds); 3) *Tension*: The temporary energy rise outlives the adjustment-process by several seconds. This is the critical phase of heightened tension within the technical system when malfunctions can occur (duration: 8 seconds); 4) *Relaxation*: The tension is removed from the production system as the energy level resets down to normal (duration: 5 seconds); 5) *Pause*: Phase before next interval starts (duration: 2 seconds).

Rules. The team task is to watch out for symptoms of malfunctioning. For the participants knowledge about constraints in the system dynamics mitigates the uncertainty about the location of malfunction occurence. The following rules apply: I.) a malfunction is (and only is) indicated by a setpoint kickback; II.) a setpoint kickback only occurs after an event of automatic adaption; III.) there are two necessary preconditions for a malfunction to occur: a) a malfunction can only be expected at sites where only one event of adaptive automation took place; b) The automated adjustment must be accompanied by an abnormal high energy demand of the automation control process; IV.) a setpoint kickback will be observable for all team members at exactly the gauge where the one team member observed the setpoint change; V.) a malfunction only occurs exactly during a phase of high tension within the production-system, when the output-quantities match the new presets but energy demand level is still high (critical interval; duration: 8 seconds); VI) when the energy level resets to normal a malfunction no longer is to be expected.

Task steps. So with the beginning of each interval the task-steps for the team are to: A) exchange information about set-point changes and collectively narrowing down at which site the first precondition for a malfunction is given; B) collectively observing, if the following adaptation process is accompanied by an abnormal high energy demand; C) distribute the information of the gauge position where the setpoint-change was observed so all team members can build up an expectation and closely monitor the relevant gauge during the critical interval; D) In case of a setpoint kickback perception: Respond with pressing the “diagnosis” button within 4 seconds. A scenario consists of N=144 intervals of 30 seconds durations, with 89 intervals showing relevant events, from that six intervals showing a malfunction event.

Research Question

The purpose of the reported exploratory experiment is to have a manipulation check whether a collective state of expectation can be induced by interactions related with a monitoring-task, and if this can be showed by team gaze indicators. Barzantny et al. (2017) provide indication that individual differences in frequencies of fixations on the relevant AoI during phases of expectancy discriminate between higher and lower performers in the malfunction detection task. It is an open question if these results can be replicated for patterns of gaze-behavior on the team level.

Material and Setup

A team of three participants took part in each experimental session. They were located within the same room, each one sitting in front of a 60Hz 1920 x 1080 Pixel 21‘‘LCD display. During the session mobile dividing walls to visually separate the working positions. The eye movements were recorded with the Remote Eye Follower System from LC Technologies, Inc. The system worked at 120Hz and was linked to the simulation ConCenT using a common time stamp. The raw data were processed using the Nyan software. The fixation detection algorithm has been set to a certain point on the screen with a minimum threshold of six eye movements with a deviation threshold of 25 pixels. All successive fixations that fell into an area of interest (AOI) were categorized as "dwell time".

Participants

Among the 63 participants were 41 applicants for a air traffic controller trainingat DFS (Deutsche Flugsicherung GmbH), the others were students from local universities. The participants were between 18 and 34 years of age (M = 21.57, SD = 3.39). 47.6% were females, 52.4% were male. The participants received 25 to 35 € as compensation for the two to three-hour experiment.

Experimental Session

After filling out a demographic questionnaire the participants received a written instruction. After performing a guided tutorial they performed the experimental scenario in one team trial of 72 inutes duration. A session ended with a task-related questionnaire.

Measures

During the critical phase of uncertainty there are three types of AoIs that are relevant for successful task performance: a) process gauges where the rules apply so that setpoint-kickbacks can be expected as an AoI for gaze measures indicating the close monitoring of the critical task element; b) the energy gauge as an AoI for gaze measures indicating the participants’ expectations for the offset of the critical phase; c) the response button to report setpoint-kickback as an AoI indicating action preparation. This report focuses on gaze measures related to these situational relevant AoI. The team was chosen as the entity of analysis. Only gaze data emitted during the critical phases were analyzed. For the purpose of this exploratoy analysis gaze-data registered for 30 intervals were prepared. The criterion for their selection was: all intervals showing a malfunction (N=6) plus the preceding four intervals in each case. All intervals were then excluded from the analysis that featured no critical event (N= 10), or more than one critical event (N= 9), so 11 of the prepared intervals remained for data analysis. The 11 intervals that were analyzed are representants of six exemplary temporal sections, covering points in scenario time from 0h 0min. 12sec. to 1h 2min.30sec.. Fixation counts and dwell times were analysed for the class of relevant AoI. This was constituted by all interactive elements on the screen, i.e. displays of relevant information (energy-demand and production-output) and the input element (response task) relevant for task performance at a given critical phase. After data clearing 18 complete teams with N=54 participants datasets were considered for the analysis. Individual data were aggregated over teams. The percentage of mean dwell times of a team on the relevant elements (3 AoIs) during critical intervals were calculated. Also the mean of the individual fixation frequencies measured for the relevant elements (3 AoIs) during each critical intervals were calculated for each team.

Results

Malfunction identification rate was generally high and increased with the third event in sequence (Events 1: 48.1%; 2: 48.1%; 3: 84.6%; 4: 80.8%; 5: 80.8; 6: 90.4%).

Posthoc Grouping of Teams

Teams were split into posthoc groups according to their collective fixation frequency measured for the relevant AoI during the critical phases (above or below mean), with higher values reflecting adequate resource allocation, and the standard deviation of this measure within a team (above or below mean) with low values reflecting the homogenity of resource allocation within a team. The rationale behind is that a recurrence in gazes between team members should result in relative low standad errors of fixation frequencies, may frequency be on a higher or lower level. These standard errors could serve as a metric for patterns of gaze-behavior, but without reflecing sequence information. This resulted in four categories with teams that were a) homogeneous in fixation frequencies for relevant objects on a high level (N=3), heterogenious in fixation frequencies for relevant objects on a high level (N=5), Homogeneous in fixation frequencies for relevant objects on a low level (N=7) and Heterogeneous in fixation frequencies for relevant objects on a low level (N=3). Class of teams with homogeneous high fixation counts represents only teams with at least two participants that show a malfunction identification rate above mean.

Team Monitoring Performance

A 19 (Interval in sequence) x 3 (Relevant AoI: Relevant Power-Gauge or Relevant Output-Gauge or Relevant Response Button) ANOVA with repeated measures with gaze-based posthoc team classification as a between subjects variable was performed for percentage of dwell times registered for a team during the interval of uncertainty.

During critical phases team ressource allocated to Relevant AoI accroding to the *Interval Position in Sequence* changes over the scenario (Sum of Means for Intervals 1-11 = 37,73%, 40,14%, 41,61% [*Section 1*] 66,35%, 54,27% [*Section 2*], 51,42%, 52,11% [*Section 3*], 67,29% [*Section 4*], 75,92%, 49,82% [*Section 5*], and 64,03% [*Section 6*] respectively; $F(5, 10) = 8.195$, $p < .05$; $\eta^2 = 0.94$). In the course of the scenario the amount of attentional resources teams allocated to relevant task elements during critical phases increased (polynomial contrast for a linear trend: $F=40.35$, $df= 1$ $p < .0001$, $\eta^2=0.74$). However, there were no *Interval Position in Sequence* by *Team Classification* [$F (21, 30) = 1.68$, $p < .11$, $\eta^2=0.71$] interactions (cf. Figure 2), indicating that team-classes do not differ in their trend to allocate more attentional resources to relevant AoIs over time.

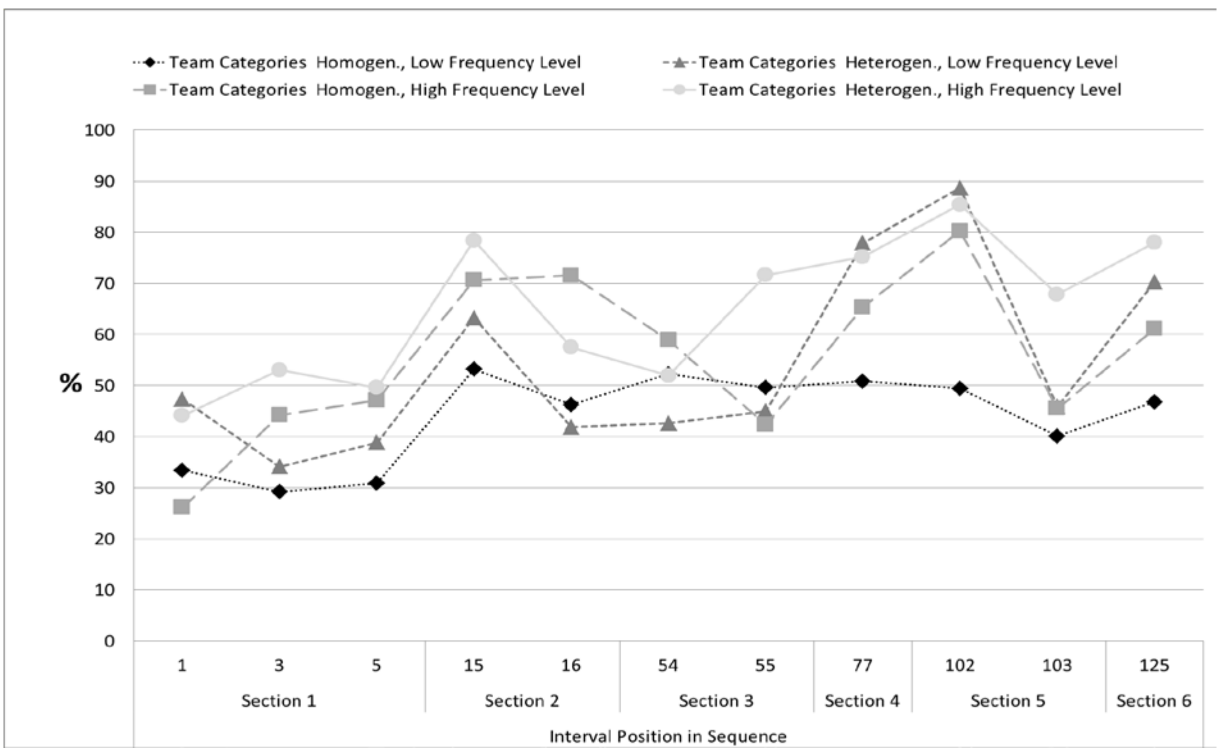

Figure 2: : Mean percentage of teams-dwell times on the relevant elements during critical intervals: *Position of Critical Interval in Seqence* by *Team Categories*

The allocation of resource of attentional resources to focussed monitoring also differed between the three relevant AoI (Means for AoI 1-3: 12.79% *[Power-Gauge], 33.85% [Output-Gauge] ,7.97% [Response Button]*); $F(2, 13) = 60.51$, $p < .0001$; $\eta^2= 0,90$) [in sum, more attention was allocated to relevant AoIs than to non-relevant AoI, which part of the 44.61% residual data can be assigned to non-relevant AoI, supposedly reflecting the allocation of attentional resources for distributed monitoring]. Tests of innersubject-effects indicate a *Relevant AoI* by *Interval Position in Sequence* interaction, F=3.04, df= 16.08 p < .0001, η^2=0.18, showing that teams were also changing their ressource allocation between the relevant AoI over time (cf. *Figure 2*). However there were no *Relevant AoI* by *Team Classification* [F(6, 28) = 1.29, p = .30, η^2=0.22] (cf. Figure 2), interactions, indicating that team-classes do not differ in their the way of monitoring relevant elements.

Discussion & Conclusions

The findings of the current study provide evidence that the quality of interactions, operationalized by measuring the prediction of correct locations using gaze data analysis, and the patterns in collective states of expectations can be induced by the chosen experimental manipulation. There was a trend that teams allocated more resources for providing focal attention and less resources for providing distributed attentions in the course of the scenario. This may be attributable to an increase in understanding of the task dynamics, paired with an increase in trust over time that events are predictable (decreases in the focal attention observable to the end of a scenario might be attributable to fatigue). Teams also were changing their ressource allocation between the relevant AoI over time. However, the chosen method for post-hoc grouping of teams based on fixation- frequency, and its within-group standar errors might not have led to a significant discrimmination between teams with different cognitive states. Although the high effect size observed for the no *Interval Position in Sequence* by *Team Classification* interaction might fuel the assumption that insignificant results were attributable to the small sample size. Nevertheless, the use of standard errors of fixation frequencies might be a too simplistic indicator for patterns of gaze behavior on team level, since this parameter also does not account for sequence information. Therfore a next step to search for team characterising patterns would be to apply the more complex method of gaze-recurrence analysis to data within teams of three. Furthermore, applying an design with experimental grous is the way to proceed. Especially as integrated gaze data analysis promises to provide an advanced method to test for Cooke et al`s (2013) assumption that team cognition patterns are more agile than individual performance patterns. Maybe gaze data will continue to tell this story, for the best of future ATM design.

References

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In Individual and Group Decision Making: Current Issues (pp. 221-245). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. Cognitive Science, 37, 255-285.

Cooke, N. J. & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. Journal of Cognitive Engineering and Decision Making, 3, 27-46.

Dekker, S. W. A. & Woods, D. D. (1999). To intervene or not to intervene: the dilemma of management by exception. Cognition, Technology & Work, 1, 86-96.

Hauland, G. (2008). Measuring individual and team situation awareness during planning tasks in training of en route air traffic control. The International Journal of Aviation Psychology, 18, 290-304.

Hess, S. M., MacMillan, J., Serfaty, D., & Elliott, L. (2005). From cognitive task analysis to simulation: Developing a synthetic team task for AWACS weapons directors. DTIC Document.

Schulze Kissing, D. & Eissfeldt, H. (2015). ConCenT: Eine Simulationsplattform zur Untersuchung kollaborativer Entscheidungsprozesse in Leitzentralen. In M. Grandt & S. Schmerwitz (Eds.), 57. Fachausschusssitzung Anthropotechnik: Kooperation und kooperative Systeme in der Fahrzeug- und Prozessführung (pp. 157-170). Bonn: Deutsche Gesellschaft für Luft- und Raumfahrt - Lilienthal-Oberth e.V.

Schulze-Kissing, D. & Bruder, C. (2016). Der Einsatz synthetischer Aufgabenumgebungen zur Untersuchung kollaborativer Prozesse in Leitzentralen am Beispiel der „Generic Control Center Task Environnment“ (ConCenT). Proceedings of the Workshop Kognitive Systeme 2016, Bochum. Retrieved from http://elib.dlr.de/109031/1/63_215_1_SM.pdf