CHAPTER FIFTEEN

Measuring Situation Awareness through Automated Communication Analysis

Peter W. Foltz\(^1\), Cheryl A. Bolstad\(^2\), Haydee M. Cuevas\(^2\),

Marita Franzke\(^1\), Mark Rosenstein\(^1\), & Anthony M. Costello\(^2\)

Pearson Knowledge Technologies\(^1\) and SA Technologies\(^2\)

Acknowledgements

Work on this paper was partially supported by a Phase II Small Business Innovative Research Contract (N00014-05-C-0438) awarded to Cheryl Bolstad and Peter Foltz from the Office of Naval Research (ONR). The views and conclusions contained herein, however, are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR, U. S. Navy, Department of Defense, U. S. Government, or the organizations with which the authors are affiliated. We thank an anonymous reviewer for helpful suggestions on improving this chapter.

Introduction

As the military undergoes significant changes including moving toward smaller, more deployable, dispersed forces, the need to find new methods to analyze and assess team performance has increased significantly. In this new modernized military, warfighters operate in a distributed fashion, sharing ever increasing volumes of information, making it more difficult to establish the high degree of shared understanding and situation awareness (SA) that is needed to function effectively. Moreover, prior research has
established SA as a promising method to predict team performance, and as an insightful theory to analyze when and why team performance might break down (Endsley 1995). Consequently, valid and reliable assessment of the team’s SA is critical for evaluating team performance and diagnosing performance deficits.

To address this need, in this chapter, we describe our approach to assessing team performance. This primarily involves first observing team actions and communications and then linking these observations to team performance directly or to surrogates and/or components of team performance, such as decision making and SA. While this limits our explanatory power to those aspects of performance being modeled (such as SA or certain communication events), in many cases of training or operational settings, this level of detail is sufficient. For instance, in a training situation, this approach can be used to notify an instructor that one team is using communication that is off topic or that another team has lost SA, or in a cockpit situation, notify the flight crew that the team has not discussed an important checklist item.

We begin this chapter with a theoretical overview of the SA construct, distinguishing among the different levels and types of SA. We then review the theoretical and empirical research underlying our measures of SA, focusing on the application of the TeamPrints methodology to analyze team communications. The emphasis of this chapter is to present recent promising research that suggests how team communication analytical techniques can be combined to develop an automated online analysis of team performance. We conclude with implications for the design of tools for automated monitoring and training.
Situation Awareness and Team Performance

Numerous studies have highlighted the vital role of SA to ensure successful performance in complex operational environments (e.g., Artman 2000; Endsley 1993; Furniss and Blandford 2006). SA can be defined as “…the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” (Endsley 1995, p. 36). Building SA, therefore, involves perceiving critical factors in the environment (Level 1 SA), comprehending what those factors mean, particularly when integrated together in relation to the individual's goals (Level 2 SA), and, at the highest level, projecting what will happen in the near future (Level 3 SA). Although alone it cannot guarantee successful decision making, SA does support the necessary input processes (e.g., cue recognition, situation assessment, prediction) upon which good decisions are based (Artman 2000). The higher levels of SA are especially critical for timely, effective decision making.

At the group level, SA has been investigated in terms of team SA and shared SA. Team SA can be defined as “the degree to which every team member possesses the SA required for his or her responsibilities” (Endsley 1995). Implicit in this statement is the idea of consistency and synchronicity between individual team member’s SA. The success or failure of a team depends on the success or failure of each of its team members. In contrast, shared SA refers to the "degree to which team members possess the same SA on shared SA requirements” (Endsley and Jones 1997). Ideally, each team member shares a mutual understanding of what is happening on those SA elements that are in common. Unlike team SA, shared SA requirements exist as a function of the essential interdependency of the team members.
Considerable research has documented the importance of SA for effective teamwork across a broad range of domains, from aviation (e.g., Hartel, Smith, and Prince 1991; Nullmeyer, Stella, Montijo, and Harden 2005) to emergency medical dispatch centers (e.g., Blandford and Wong 2004; Rhodenizer, Pharmer, Bowers, and Cuevas 2000) to military command and control (e.g., Gorman, Cooke, and Winner 2006; Kaempf, Klein, Thordsen, and Wolf 1996). For example, in their study of CRM behaviors in military MC-130P crews, Nullmeyer and Spiker (2003) found that SA stood out as the Crew Resource Management behavior most strongly associated with mission planning and that crews receiving high SA ratings often also demonstrated good mission performance.

Our approach to linking SA with team performance is consistent with the structural model of team collaboration described earlier in this volume (see chapter 2). The model was created to aid in understanding the relationships between cognitive mechanisms and team collaborative problem solving abilities. It provides a framework from which further research can be performed to assess the underlying processes of team performance, such as SA or communication. The model delineates four main collaboration stages: 1) team knowledge base construction, 2) collaborative team problem solving, 3) team consensus, and 4) outcome evaluation and revision. SA is used throughout each stage to construct a shared understanding of the situation at hand and to decide on the appropriate course of action. During each of these stages, individual, team, or shared SA will play a more prominent role depending upon the task being performed. For example, during the first stage, team knowledge base construction, teams strive to gather the necessary information to form an assessment of the situation (Level 1 SA).
Their assessments are then shared with one another to form a more complete common operating picture (shared SA). This shared SA may then be used by the team to solve the problem at hand. Thus, their performance may be based on the accuracy of their shared and team SA.

**Team Communication and Situation Awareness**

Communication is an essential mechanism for achieving the critical cognitive processes at each of the four main stages of the team collaboration model (see chapter 2, this volume). Independent of operational context, team communication, either in the form of verbal transmissions or electronic text exchange (as in e-mail or chat) is central to the ongoing collaboration of any team. Communication is necessary to share individual cognitive representations, including developing SA, building a joint knowledge base, solving problems, forming consensus, evaluating outcomes, and revising plans and actions. As such, communication is indicative of the individual and joint cognitive processes that underlie the stages of team decision making.

Notably, research has shown that communication increases team effectiveness by helping teams form shared SA and shared mental models (e.g., Brannick, Roach, and Salas 1993; MacMillan, Entin, and Serfaty 2004). Team communication (particularly verbal communication) supports the knowledge building and information processing that leads to SA construction (Endsley and Jones 1997). Thus, since SA may be distributed via communication, utilizing machine learning, models can be created that draw on the verbal expressions of the team as an input and provide performance estimates as an output. The logical next step is to attempt to develop technology that would make real-
time automatic analysis of team communication with respect to SA and team performance possible. To achieve this goal, two promising methods of predicting team performance were joined to enhance each other’s predictive and inferential power. Specifically, we combined the explanatory capacity of the SA construct with the predictive and computational power of the TeamPrints methodology. Next, we discuss the theoretical background and empirical support for TeamPrints.

**TeamPrints: LSA-based Analysis of Real-Time Communication**

TeamPrints is a system that uses computational linguistics and machine learning techniques coupled tightly with Latent Semantic Analysis (LSA) to analyze team communications. LSA is a fully automatic technology for modeling and matching discourse content. This technique’s special capabilities for communications analysis include its ability to represent the entire conceptual content of verbal communication rather than surrogates such as keyword, titles, abstracts, or overlap counts of literal words. In addition, LSA’s representation of the similarity of two words, sentences, utterances, passages, or documents closely simulates human judgments of the overall similarity of their meanings. LSA also represents two passages on the same topic, but phrased in different vocabulary as similar.

LSA presumes that the overall semantic content of a passage, such as a paragraph, abstract or coherent document, can be closely approximated as a sum of the meaning of its words:

\[ \text{Meaning of paragraph} \sim \text{meaning of word}_1 + \ldots + \text{meaning of word}_n \]
Mutually consistent meaning representations for words and passages can thus be derived from a large text corpus by treating each passage as a linear equation and the corpus as a system of simultaneous equations. In standard LSA, the solution of such a system is accomplished by the matrix decomposition Singular Value Decomposition (for details on the theoretical and mathematical underpinnings of LSA, see Landauer, Foltz, and Laham 1998). LSA’s effectiveness in simulating similarity of meaning for humans has been empirically demonstrated in many ways. For example, by matching documents with similar meanings but different words, LSA improves recall in information retrieval, usually achieving 10-30% better performance cetera paribus by standard metrics (Dumais 1994). After training on corpora from which humans learned or might have, LSA-based simulations have passed multiple choice vocabulary tests and textbook-based final exams at student levels (Landauer et al. 1998). LSA has been found to measure coherence of text in such a way as to predict human comprehension as well as sophisticated psycholinguistic analysis, while measures of surface word overlap fail badly (Foltz, Kintsch, and Landauer 1998). By comparing contents, LSA predicted human ratings of the adequacy of content in expository test essays nearly as well as the scores of two human experts predicted each other, as measured by ~90% as high mutual information between LSA and human scores as between two sets of human scores.

Typically, LSA ignores word order within documents. However, in much of the present work, additional statistical Natural Language Processing (NLP) techniques are used in conjunction with LSA to account for the syntactic and grammatical differences found in sentences (Foltz and Martin 2004). This combination of LSA and NLP techniques to model team communication is what we have termed TeamPrints.
Predicting Team Performance

TeamPrints has been evaluated favorably in terms of its ability to predict team performance. For instance, as a proof of concept, TeamPrints was able to successfully predict team performance in a simulated unmanned aerial vehicle (UAV) task environment based only on communications transcripts (Foltz and Martin 2004; Gorman, Foltz, Kiekel, Martin, and Cooke 2003). Using human transcriptions of 67 team missions in a UAV environment, TeamPrints reliably predicted objective team performance scores (LSA alone, \( r = 0.74 \); LSA combined with additional computational linguistic analysis measures, \( r = 0.79 \)).

In a similar study, Foltz, Martin, Abdelali, Rostenstein, and Oberbreckling (2006) modeled Naval air warfare team performance in simulated missions consisting of 8 six-person teams. Sixty-four transcribed missions were obtained from the Navy Tactical Decision Making Under Stress (TADMUS) dataset collected at the Surface Warfare Officer’s School (see Johnston, Poirier, and Smith-Jentsch 1998). In this scenario, a ship’s air defense warfare team performed a detect-to-engage sequence on aircraft in the vicinity of the battle group, and reported it to the tactical action officer and bridge. Associated with the missions were 16 subject matter experts generated performance measures, including elements associated with SA, such as the passing of information, seeking of information, and updating the team as to the situation. All predictive models for each of these 16 measures were highly significant with correlations to performance measures ranging from \( r = .45 \) to .78, demonstrating that the models can provide accurate
predictions for a range of different independently-rated team performance scores, including communication quality, SA, coordination, and leadership.

*Predicting Situation Awareness*

Having established the efficiency and effectiveness of the TeamPrints methodology, we next focused on applying this approach to predict SA. As discussed earlier in this chapter, communication is a critical component of the process of team decision making as a whole, and of the component processes of developing SA, building a joint knowledge base, problem solving, consensus building, and outcome evaluation. These components should be extractable and quantifiable from recorded streams of team communication. In particular, we propose that individual and team SA is latent in team communications and, thus, should be measurable through an analysis of the communication stream. The research previously described in this chapter provides support for the utility of TeamPrints for reliably predicting objectively measured task performance as well as measures of SA. The objectives of the current research effort, therefore, were to (1) replicate these findings using new tasks, (2) focus on using TeamPrints to predict SA, and (3) attempt to build a TeamPrints-based automatic tagging method that would parse communication streams into larger groupings and automatically assign SA levels to these communication units.

In order to achieve these objectives, we followed a two-pronged approach. First, we identified an existing data set, the Noncombatant Evacuation Operation (NEO) Mission Scenario (collected as part of the Collaborative Knowledge in Asynchronous Collaboration (CASC) Phase II Project; Warner, Wroblewski, and Shuck 2003), to help
us define our analyses methods. With the knowledge gained through these exploratory analyses, we then created an experimental plan using the C$^3$Fire simulation package that allowed us to collect team communications and SA and task performance data in a more optimal way to assure that the necessary analyses could be conducted. Each of these efforts will be described next.

**Exploratory Analyses: The NEO Mission Scenario Data Set**

Our exploratory analysis of the NEO data set has been published elsewhere (Bolstad, Foltz, Franzke, Cuevas, Rosenstein and Costello, forthcoming), and, thus, will only be summarized here to highlight the application of the TeamPrints methodology. The NEO Mission Scenario asks participants, as part of a three-person team, to develop a plan to rescue stranded personnel from a church basement on a remote island in the middle of guerilla warfare (for a detailed description of the original study, see Warner et al. 2003). Our analyses of this data set focused on the written and recorded communications of the teams while creating a plan, and a set of independently derived performance scores. We evaluated how well our analysis of the communication data using TeamPrints would successfully predict objective measures of team performance and SA level.

*Predicting Performance Using TeamPrints*

Each team’s performance on the NEO Mission Scenario task was scored by independent raters on a 100-point scale. The average score across all 32 teams was 83.8 (SD = 7.2). Using variables based on LSA nearest neighbors (transcripts that were near in semantic space), we were able to predict team performance of 16 teams with a jackknife
correlation of .77 (for a description of the jackknife technique, see Yu 2003). The model with the best fit to the data used four variables, all measuring different aspects of the similarity of the team communications of the predicted team to successful teams. As in earlier work (e.g., Foltz and Martin 2004), the closer (more similar) a transcript was to its nearest neighbor the better the team’s performance, how well a transcript’s nearest neighbors performed was a positive indicator of how well this team did, and the more variability there was in the performance of the nearest neighbors, the worse a team was likely to perform.

We also evaluated whether the communication data alone would predict the team’s experimental condition (static vs. dynamic). In dynamic conditions, teams received new information half-way into their planning process. Intuitively, the communication of teams in this condition should be different, because they had to adjust their joint knowledge and decision making based on this information. As hypothesized, TeamPrints predicted assignment to static or dynamic condition with 75% accuracy using a discriminant analysis. Indeed, our TeamPrints analysis of the communication data made more accurate predictions of condition than the team performance scores, which had been assigned by human raters (56% accuracy).

In summary, we achieved the first goal of our plan, that is, we replicated our earlier findings that TeamPrints analyses of the communication data stream allows for reliable prediction of team performance. We also extended these initial findings to show that our analysis was sensitive enough to detect systematic differences between conditions.
Predicting SA Level Using TeamPrints

Because the NEO data set was collected for related but different purposes, no traditional online or post-experimental measures indicating actual or perceived SA were administered. Accordingly, an alternative method of assessing SA post-hoc was utilized (for further details, see Bolstad et al. forthcoming). Seven of the team communication transcripts were analyzed manually by two human raters. The raters first grouped communication exchanges into meaningful, coherent units and then tagged each grouping by SA level (SA Level 1, 2, or 3). SA level was then predicted in a series of hold-one-out experiments, where a model was created by training TeamPrints using 6 of the 7 transcripts, and the SA levels of the held-out transcript were predicted.

The average exact agreement obtained across this set of experiments was 51.1%, which is less reliable than TeamPrints has been in other applications (see, for example, the analysis of the TADMUS data set, described earlier in this chapter). Still, this estimate is better than a pure random assignment, which would have produced an exact agreement of only 33.33%. Further, the robustness of the TeamPrints model may have been constrained by: first, the relatively small training set of only 7 transcripts; and second, having to infer SA based on post-hoc analysis of team communications, without real-time objective SA measures from the actual experiment for comparison. Given the preliminary nature of this exploratory analysis, these findings are promising nonetheless.

Predicting Team Performance Using SA Levels

The last step in this exploratory series of analyses was to attempt to predict team performance using the human and machine-produced SA-level tags. Teams with higher
levels of SA would be expected to show better performance. However, analyses did not reveal any significant correlations between the different levels of SA and team performance, neither for the human- nor the machine-derived tags. Nevertheless, the correlations between the SA levels and performance were in the predicted directions, where higher proportions of Level 1 SA tended to be correlated with lower performance, and higher proportions of Level 2 and 3 SA (indicative of higher levels of data processing) tended to be correlated with higher levels of performance.

Summary of NEO Data Exploratory Analyses

Through our exploratory analyses of the NEO data set, we were able to provide evidence to support that analyses of communication streams using TeamPrints could reliably predict team performance as well as detect systematic differences between conditions. We also found some evidence to suggest that analysis of team communications using TeamPrints can be used to identify SA levels among teams. Finally, although the correlations between the different SA levels and team performance were not significant, our analysis demonstrated that it is possible to use TeamPrints to tag SA level automatically, and that these automatically derived tags should, in principle, predict performance similar to human tags. Clearly, we would expect more convincing results in an experimental situation that allowed measuring SA level with more proven methods. The experiment described next was designed to create this type of data set.

C3Fire Simulation Experiment
The computer game C3Fire (http://www.c3fire.org) was selected as the test bed for investigating team collaboration, communication, and SA. C3Fire was developed as a dynamic team simulation used for team training and research of command, control and communication (C3) in a collaborative environment. The C3Fire computer-based system generates a task environment possessing complex, dynamic and opaque characteristics, similar to cognitive tasks normally encountered in real-life systems.

**Experimental Design**

The study employed a one-factor within-groups design, with scenario complexity serving as the independent variable. Dependent variables include assessment of participants’ communication exchanges, SA, and task performance derived from the simulation data.

**Participants**

Twenty-four students (male = 23; female = 1; mean age = 32.42) from the Naval Postgraduate School, Monterey, CA, voluntarily participated in this experiment for course credit. Participants were randomly assigned to 3-person teams, creating a total of 8 teams. Treatment of all participants was in accordance with the ethical standards of the American Psychological Association.

**C3Fire Scenarios**

The C³Fire scenarios created for this study involved a 3-person team task played using standard desktop PC-based computers. C3Fire simulates the outbreak of a fire and the team’s primary goal was to put the fire out, saving as many houses and as much terrain as
possible. This required participants to coordinate their efforts to control the fire through the use of fire, fuel, and water trucks. Two participants were each assigned the role of Fire Chief and were responsible for controlling a varying number of fire trucks. A third participant was assigned the role of Water / Fuel Chief and was responsible for controlling a varying number of water and fuel trucks. The objective of the C3Fire game was to keep the map area clear of fire by closing out any existing cells (squares) on fire and preventing any further spread to other cells. The shading of the cells on the map indicated the status of the fire, which included one of four states:

- Clear (no shading) – area is clear of fire
- Fire (red) – area is burning (on fire)
- Burned Out (black) – area is no longer burning; fire burned itself out
- Closed Out (brown) – area is no longer burning; fire successfully put out

Participants were geographically distributed in different rooms and could only communicate with each other using USB headsets and Voice-Over Internet Protocol (VOIP) software. Participants viewed the fire via networked interfaces and controlled their trucks via the GUI. All participants could see the map showing the developing fire and the assets that needed to be protected. However, each participant could only see his/her own trucks and their status indicators (fill level). Their team members’ trucks were only visible when these were in the 9-square radius on the map surrounding the participant’s own trucks. This set-up made it necessary for participants to verbally communicate their positions, fill-levels, and their (intended) actions to coordinate their plans for fighting the fire.
Six scenarios were created varying in complexity to elicit targeted communication and coordination team behaviors and create opportunities to assess SA. Scenario complexity was manipulated by varying the number of trucks assigned to each participant and other C3Fire parameters, such as wind speed, number and placement of assets (e.g., houses to be protected), initial fill level of the trucks, and burn rate of the fires. Increasing scenario complexity inherently heightened the necessity for more frequent and succinct communications as the scenarios became more difficult.

Objective SA Measure: SAGAT Queries

Using the Situation Awareness Global Assessment Technique (SAGAT) (see Endsley 2000), 13 queries were created to assess participants’ SA during each scenario. These queries were derived from a Goal Directed Task Analysis (GDTA) (for a description on GDTA creation, see Endsley, Bolte, and Jones 2003) of the C3Fire domain. Seven queries assessed participants’ Level 1 SA (perception), such as, for example, “Indicate the province to which Fire Truck 3 (F3) is most closely located.” Six queries assessed participants’ Level 2 and 3 SA (comprehension and projection), such as, for example, “Indicate which province is most threatened by the fire.” The SAGAT queries were presented to each participant on their computer screen once during each scenario, with presentation order and time varied randomly across trials. Participants’ responses were compared to ‘ground truth,’ as recorded by the computer, to provide an objective evaluation of their SA. Team-level SA scores were calculated by averaging across individual team members.
Subjective SA Measure: PSAQ

Participants’ subjective SA was assessed by administering the Post-Trial Participant Subjective Situation Awareness Questionnaire (PSAQ) (Strater, Endsley, Pleban, and Matthews 2001) at the end of each scenario. The PSAQ consists of three items designed to solicit self-report of participants’ perceived workload (“Rate how hard you were working during this scenario”), performance (“Rate how well you performed during this scenario”), and awareness of the evolving situation (i.e., their SA) (“Rate how aware of the evolving situation you were during the scenario”). Ratings were recorded on a 5-point scale, with responses ranging from 1 (low value) to 5 (high value). Team-level scores were calculated by averaging across individual team members.

Team Communication

Audio recordings of the verbal team communications occurring during the scenarios were transcribed and were annotated to indicate scenario start and end times as well as to note when the SAGAT queries were administered. The transcripts were then analyzed using the TeamPrints methodology (as described later in the results).

C3Fire Task Performance

Task performance data was extracted from the event logs automatically generated by the C3Fire system. Performance was measured in terms of how well the team was able to close out the fire, with minimal loss to critical objects. Specific metrics included determining the number of cells on the map that were burning (red), burned out (black),
and closed out (brown). These values were calculated twice for each team: at each SAGAT stop and at the end of each scenario.

Results

The C3Fire experiment data was analyzed using the same rationale as used with the NEO data. Specifically, team performance was predicted directly using TeamPrints alone and with the two SA measures, PSAQ and SAGAT. TeamPrints was also used to predict SA level (as measured using SAGAT). Finally, the TeamPrints model of SA was used to predict performance.

Communication protocols Each of the eight teams ran through the six scenarios. Data from one scenario was lost due to technical problems, leaving communication data from 47 transcripts. Table 15.1 shows a brief exchange illustrating the richness of the verbal data collected from these exercises.

Table 15.1

<table>
<thead>
<tr>
<th>Illustrative Example of Team Communications during a C3Fire Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water/Fuel</td>
</tr>
<tr>
<td>Fire Chief 1</td>
</tr>
<tr>
<td>Water/Fuel</td>
</tr>
<tr>
<td>Fire Chief 1</td>
</tr>
<tr>
<td>Water/Fuel</td>
</tr>
<tr>
<td>Fire Chief 1</td>
</tr>
<tr>
<td>Fire Chief 2</td>
</tr>
<tr>
<td>Water/Fuel</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>Fire Chief 2</td>
</tr>
</tbody>
</table>

These transcribed utterances from the scenarios were used in the modeling efforts. Scenario transcripts contained an average of 1217 words (SD = 429). The complete communication transcript for each scenario was used to model overall team performance and team-level PSAQ responses. To predict SAGAT scores, only the utterances up until the SAGAT stop were used. Average scenario word count at the SAGAT stop was 668 (SD = 264).

_C3Fire team performance_ As indicated earlier, the team’s primary goal was to control the fire as well as protect houses from the fire. Performance measures, therefore, focused on the total number of cells burned or burning (later termed _lost cells_) and the number of houses that were burning or burned out (_lost houses_)_. Greater values indicated worse performance. Figure 15.1 represents the total number of lost cells and lost houses at the time of the SAGAT stop. One striking aspect of this graph is the lack of variability in the number of houses lost. At the SAGAT stop, 42 out of 47 scenarios showed only 1, 2 or 3 lost houses. Given the variability in the language of the teams, this finding makes predicting performance very difficult. Thus, we focused on modeling the total number of lost cells as our main dependent variable.

[Insert Fig. 15.1 near here]

This performance measure was derived both at the time of the SAGAT stop as well as at the end of the scenario to allow meaningful modeling of the SAGAT measure.
Results showed that performance at the SAGAT stop predicted end-of-scenario performance well ($r = .88, p < .01$). If a team had a higher number of lost cells at the time of the SAGAT stop, they also tended to have higher numbers of lost cells at the end of the scenario.

As noted earlier, scenarios became progressively more difficult during the experiment. The significant Spearman rank correlation between scenario number and total lost cells ($r = .54; p < .01$) suggests that the manipulations to increase scenario complexity were effective. In general, participants appeared to find controlling the fire in later scenarios more difficult, leading to higher numbers of lost cells at the end of these scenarios (see Fig. 15.2).

[Insert Fig. 15.2 near here]

**PSAQ** Results on the analysis of team-level average for the PSAQ responses revealed an interesting pattern of results. Subjective judgments of SA and C3Fire performance level were highly correlated ($r = .72, p < .01$), but correlations between perceived workload and perceived SA ($r = -.09, p > .10$), and perceived workload and perceived performance ($r = .33, p < .05$) were not as strong. One explanation for this pattern may be that SA and performance are attributed to self, and, thus, are found to be more similar, whereas workload is attributed to the situation, and might, therefore, be judged on a different dimension. Interestingly, only perceived workload was a good predictor of actual C3Fire performance (as measured by number of lost cells) ($r = .68, p < .01$). Correlations between perceived SA and actual performance ($r = -.15, p > .10$) and
perceived and actual SA (as measured by the SAGAT scores) \( r = -.38, p < .05 \) were lower.

**SAGAT** One would expect higher SAGAT scores to be associated with a fewer number of lost cells, indicating better performance. However, although in the expected direction, the correlation between the team-level SAGAT score and C3Fire performance (total lost cells at the SAGAT stop) was not significant \( r = -.13, p > .10 \). When correlations were evaluated by query type (SA Level 1 vs. SA Level 2/3), slightly better predictions were obtained using the SA Level 2/3 queries \( r = .21, p > .10 \), but not enough for statistical significance. Thus, the modeling results were obtained at the aggregate level of average team performance across all SAGAT queries.

**TeamPrints** Having established our measures, we used TeamPrints to create a number of models. All models were variants of the hold-one-out paradigm, where a subset of the data was used in training the model, and remainder were used to test the model’s performance. Specifically, we used all but one of the scenarios (46) to predict the held-out-scenario. This type of experiment was performed 47 times, so that each scenario was predicted once, using the other scenarios as the training set. The values reported represent the correlations between the 47 predictions and 47 actual values.

Training TeamPrints on the C3Fire performance score, the held-out performance was significantly correlated with the predicted performance \( r = .73, p < .01 \), suggesting that the TeamPrints analysis of the communication data was able to predict performance quite well. However, for the TeamPrints model of the SAGAT data, the correlation
between the predicted and actual SAGAT scores was not significant \((r = .13, p > .10)\).

With regard to the correlations between predicted and actual responses on the three PSAQ items, only the model for perceived workload was significant \((r = .62, p < .01)\). No significant correlations were found for either perceived performance \((r = .22, p > .10)\) or perceived SA \((r = .22, p > .10)\).

**Summary of C3Fire Experiment**

Overall, the analysis of team communications using TeamPrints revealed several promising results. The number of lost cells was shown to be a valid and reliable objective measure of team performance. Using this as the dependent variable in a number of analyses, we found that subjective judgments of workload, as well as team communications, as analyzed through our TeamPrints models, were reliable predictors of team performance. We were also able to predict the workload measure using the TeamPrints models, suggesting, that these three constructs measure closely related team performance components.

Unfortunately, the objective measure of the team’s SA during the scenarios (SAGAT queries) did not correlate significantly with the performance measure, nor were we able to use TeamPrints to make reliable predictions about the team’s SA. It should be noted that the SAGAT is typically administered multiple times during task performance to truly capture operators’ SA in a given situation. However, in this study, SAGAT was administered only once in a single trial and, thus, does not represent participants’ SA for the scenario, but instead their SA at the time of the stop. Thus, matching a single SA score to a communication measure that is continuous may have caused these low
correlations. Nonetheless, although not significant, the positive correlation for the Level 2 and 3 SA queries suggests that queries probing higher levels of SA may yield greater predictive value of team performance. Specifically, teams scoring better on queries assessing situation comprehension and projection were also performing at a higher level. This finding is consistent with the results from our NEO data exploratory analysis, which showed that teams expressing proportionally more Level 1 SA performed worse.

**Conclusions and Future Directions**

While the SA construct is widely known, the quantification and measurement of SA is a relatively new field. A critical aspect of SA that also needs to be understood is that it is dynamic in nature and thus continually changing. Many metrics of SA do not take this into consideration and instead focus on a single SA measurement taken per team task or event. Communication measures not only have the potential to provide a real-time assessment of SA, they can also be used to provide diagnostic information on why a team is performing a certain way.

Overall, the results of our exploratory analyses and empirical investigations support the utility of the TeamPrints methodology for assessing and predicting team performance and only minimal support for assessing SA. We found evidence to suggest that analyses of team communications using TeamPrints can be used to reliably detect systematic differences between conditions as well as potentially identify SA level among teams. Finally, the models derived using TeamPrints reliably predicted perceived workload.

While automated communication analysis helps our understanding of
macrocognitive processes such as SA, this methodology also has implications for the
development of products to take communication input, analyze, and predict cognitive
measures in near real time. TeamPrints can serve as a useful diagnostic tool to
organizations trying to understand team cognition and performance. It would be
particularly beneficial to organizations in which the team members are continually
changing or its characteristics are changing, such as the experience of the team members
or the number and types of task they perform together. By providing relevant diagnostic
information on team cognition and SA, the system can provide an indication of the
impact of such changes on the resultant team cognitive performance. Additional
validation on novel data sets and different team contexts will still need to be performed in
order to assess the generalizability of the methods developed.

The integration of the technology into performance monitoring systems can
further be used to develop systems that can adapt interfaces to provide optimal team
performance. By performing real-time monitoring, assessing, diagnosing, and adjusting
feedback for teams, a host of products can be developed for improving training in
simulators as well as monitoring live-real time communication. Much of the expense in
training is due to having knowledgeable human trainers to monitor the teams and be able
to provide feedback at appropriate times. The technology described in this chapter can
alleviate some of this expense while still providing accurate and effective feedback to
teams and their trainers.

References


Figure Captions

Figure 15.1. Total lost cells and lost houses by team and scenario at SAGAT stop.

Figure 15.2. Total lost cells by scenario and team number at end of scenarios.