Coordination in homogeneous and heterogeneous teams

Othalia Larue OTHALIA.LARUE@WRIGHT.EDU
Ion Juvina ION.JUVINA@WRIGHT.EDU
Wright State University, 3640 Col. Glenn Hwy. Dayton, OH 45435 USA

Michael T. Cox MICHAEL.COX@WRIGHT.EDU
Matt Molineaux MATTHEW.MOLINEAUX@WRIGHT.EDU
Bruce Howard BRUCE.HOWARD@WRIGHT.EDU
Eric Nichols ERIC.NICHOLS@WRIGHT.EDU
Wright State Research Institute, Wright State University, Beavercreek, OH 45431 USA

Brandon Minnery BRAD@KAIROS-RESEARCH.COM
Kairos Research, LLC 611 W. Yellow Springs Fairfield Rd. Fairborn, OH 45324 USA

Abstract

We study markers of social cognition and team interaction across two different problem-solving tasks and domains: a controlled bridge design task and a highly dynamic military planning task. Team interactions are measured considering the task dynamic. We identify cross-domain indicators of coordination and connect them to performance. In the design task, we observed coordinated behavior in solution space exploration and reduction. In the planning task, we observed coordinated movement but less elaborate strategies. While coordination was observed through similarity in designs in the first task, it was a more complex phenomenon in the second where complementarity rather than similarity was found.

1. Introduction

Failure or success in team interactions can lead to very different consequences in critical situations. For example, the crash of United Airlines flight 173 in 1978 has been partially linked to failure in team communication which resulted in deficits in team cognition (NTSB, 1979). In contrast, efficient communication and decision-making led to the "miracle on the Hudson" for US Airways flight 1549 in 2009 (Gordon, Mendenhall, & O'Connor, 2012). Effective coordination is a marker of team cognition (Salas, Fiore, & Letsky, 2013; Salas & Fiore, 2004). Team coordination is the orchestration and timing of a sequence of interdependent actions to achieve goals and perform tasks (Marks, Mathieu & Zaccaro, 2001). Measures of team interaction and coordination across domains can also help to identify disruptions and crisis situations which in turn will allow for team crisis intervention (Molineaux & Cox, 2019). The coordination of actions can vary in cognitive complexity from moving in concert to making sense of a situation. Team sensemaking (Klein, Wiggins, & Dominguez, 2010) is defined as the process by which team members coordinate to explain the situation at hand. Sharing the view of Cooke et al. (2013), we see team cognition as arising in context. Team cognition emerges from individual and team factors as team members interact with their environment, the problem-solving task, and each other. Meaningful team interactions often occur in the presence of a critical change in the environment. A critical change is one which affects a future payoff negatively. As such, measuring team interactions must consider the task dynamic.

Design problems differ considerably from planning problems. Solutions for the former consist of component configurations that achieve functional goals and constraints (Chandrasekaran, 1990; Maher, Balachandran & Zhang, 1995); whereas, sequences of actions (i.e., plans) and their executions that achieve environmental goal states present solutions for the latter (Ghallab, Nau, & Traverso, 2016; Santos, Deloach, & Cox, 2006). Here we examine the design of a truss-style bridge by teams of engineering students and the execution of a hostage rescue mission by teams of ROTC students. In the bridge design experiment, teams are homogeneous. Participants have the same role and there is no hierarchy. Team members achieve the same task independently with some communication (team members help each other). In the hostage-rescue domain, team members have different specialties and are interdependent, yet they must act together.

This paper is the first step towards our end goal, which is to use the identified global metrics to design systems capable of instrumenting the group's global behavior for intervention at critical moments. In this paper, we identify metrics of social cognition as well as team coordination and interaction in two very distinct problem-solving tasks and domains. While team behavior research applies and adapts metrics of team sensemaking (Klein, Wiggins, & Dominguez, 2010) and team cognition (Cooke et al., 2013) to a specific domain, in this paper we focus on how metrics can generalize and transfer across very distinct applied cases and also how their divergence can help us characterize coordination in different types of team organizations (homogeneous and heterogeneous). Therefore, to measure how teams assess situations and make sense of them, we will identify in each domain behavioral indicators that could be used to analyze coordination and identify the global concepts in which they fit and which exist in both domains. The found indicators might also inform us on how the concept of coordination itself differs between domains. The structure of the task had a significant impact on what observables were available to study it. In the bridge design domain, observables are related to the design properties, while in the hostage rescue domain they have to do with teammates' positioning. In the bridge design domain, we build on the work from McComb, Cagan and Kotovsky (2015), which identified markers of team cognition. We use similar markers to study team cognition in context. We consider coordination measures in the presence of a critical change in the environment (problem statement changes in the bridge design domain; team reconfiguration following a death in the rescue domain). Finally, we identify the similarities and differences between coordination properties in each of the two structurally distinct problem-solving domains.

2. Related work

Shared cognition (Cannon-Bowers, Salas and Blickensderfer, 1999) is a theory inspired by information-processing models of individual cognition. According to the Shared Cognition approach, when a team is composed of individuals with high shared cognition, they develop similar expectations and approaches to problem solving which leads them to perform better as a team. Notably, it is hypothesized that sharing a common ground improves participants performance by reducing the amount of communication required in the team, making implicit coordination and non-verbal cues sufficient for effective communication. Reduced communication needs are especially useful in tasks with higher level of complexity. In order to assess the shared cognition, individual mental models are extracted (through survey, process tracing, …) and aggregated. Therefore, shared cognition is evaluated through metrics which assess the similarity of structures across individuals (not at the team level). While formalizing team cognition this way is efficient for homogeneous teams (same roles and knowledge) and knowledge-oriented tasks, this approach does not prove as adequate for evolving environments, and diverse heterogeneous teams. As cognitive complexity of the task increases, individuals can't maintain complete awareness of the task, and, maintaining a shared mental model of the task with a common team perspective is no longer possible. In heterogeneous teams, success relies less on sharing the same complex team mental model and more on efficient interaction in the decision making or problem-solving process.

In contrast with the Shared Cognition view, according to the Interactive Team Cognition (ITC) theory (Cooke et al., 2013; Cooke, 2015), cognition should be studied at the team-level. Shared cognition can be, in some cases, an indirect measure of team cognition. However, team cognition is not just an aggregate of cognitions of individuals from the team, it emerges from the flow of interaction of its members and provides a context for individual behavior. ITC provides a framework for the study of cognition in teams which work on task that can be handled without extensive knowledge aggregation or an extensive common state of knowledge between teammates (e.g., teammates with different roles). Influenced by ecological psychology and situated cognition, ITC describes team cognition as context dependent and unfolding in real-time. ITC is articulated around three postulates: (1) team cognition is an active process (not a static product) (2) team cognition exists at the team level (not a simple aggregate of individual knowledge) (3) team cognition is contextual. According to ITC, teams perform the following cognitive tasks: detecting relevant information, decision-making, problem-solving, remembering relevant information, planning, knowledge acquisition and solution design. While measures of shared cognition tend to rely on more traditional psychometrical measures aimed at eliciting members knowledge (measure of shared knowledge through surveys, measures of shared mental model) and then aggregating this to generate what is the shared representation, ITC uses interaction based-measures of coordination and team situation awareness (i.e., timely and adaptive response to an event) (Cooke & Gorman, 2009). For team situation awareness to happen, there needs to be: (1) a joint detection of the change from at least two team members (2) a coordinated interpretation of the change (e.g., observable through communication) and (3) a coordinated action in reaction to that change by one or more team members. Team cognition is rooted in team interaction and observed in communication and coordination patterns. Team situation awareness in ITC is measured by team interaction in the face of a critical change in the environment. A critical change is an unexpected change that poses a threat to potential future outcomes. Coordination, more specifically, is observed through information passing between teammates at significant times in the scenario (i.e., in the synthetic task environment scenario presented by Cooke (2013), the coordination metrics are the sequence and timing of identified coordination events and the communication flow (i.e., who talked to whom)). Other suggested metrics relate to coordination flexibility and stability in response to unexpected task and environmental changes.

Team sensemaking (Klein, Wiggins & Dominguez, 2010) also aims at capturing team cognition through more information than individual behaviors. Team sensemaking is the macrocognitive process by which a team adapts and coordinates in order to explain a current situation (usually under uncertain situations). While both ITC and team sensemaking can be applied to different types of environment (e.g., controlled conditions in laboratory, exercises or real-life situation), they are particularly suited for the study of team cognition in real-life environments. The nature of sensemaking emerges at the team level, defined by team member skills and contextual demands (i.e., demands of the situation). Example of independent variables that can be exploited for the study of team sensemaking are: situation novelty and uncertainty; structure of the team; experience of the team (at an individual and joint level); and performance (time or accuracy). Furthermore, Klein et al. (2010) describe a data/frame model of sensemaking where sensemaking retrospectively identifies a frame based on data and reciprocally applies that frame to data. Sensemaking ends when the appropriate frame has been identified. Through a cycle, frames provide an explanation of past events and framing of data allows for prediction of future events and subsequent re-framing based on incoming data. The strategies which can be identified as metrics of team sensemaking are frame selection, frame questioning, and replacement or selection of a new frame.

3. Two Task Domains

We obtained datasets from previous studies of two different task domains. The first one, referred to as the bridge design task (McComb, Cagan, & Kotovsky, 2015; 2018), requires coordination between team members on an optimization problem. The other one, hostage rescue mission, requires for each member to coordinate with other team members by adopting a distinct individual behavior which is required to reach the team objective (i.e., each team member has a role).

3.1 Bridge Design Domain

In this study (McComb, Cagan, & Kotovsky, 2015; McComb, Cagan & Kotovsky, 2018), sixteen teams of three participants were assigned a bridge structure (bridge spanning a chasm) design problem whose requirements are changed during the design process. Forty-eight engineering student participants were randomly assigned to the sixteen teams of three students each. All teams were assigned the same problem, but the problem statement changed twice during the experiment. The changes occurred at the same time for each team. Changes were unexpected and required participants to adapt. Figure 1 illustrates the kind of design problems teams encountered.

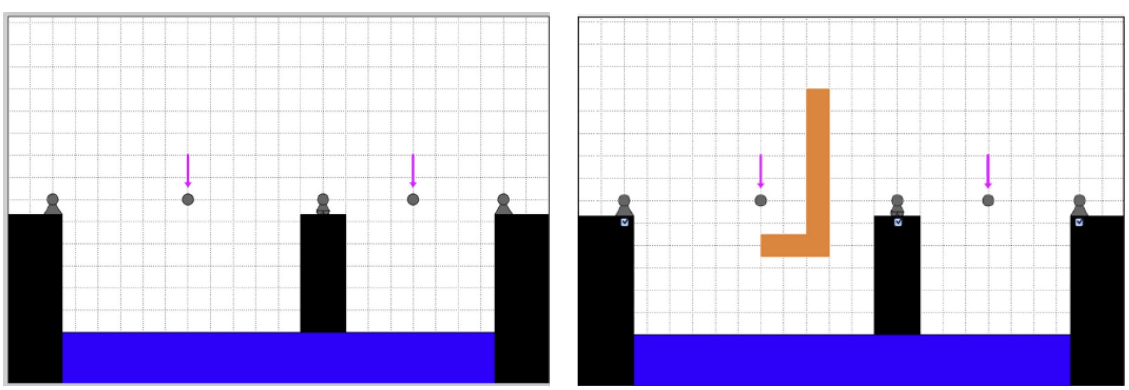

Figure 1. The bridge design domain: design of a bridge which spans a river with loads, mass and factor of safety constraints. Pink arrows point to the middle of bridge spans. In the third phase (goal 3), participants had to ensure their bridge did not overlap the orange region (second panel).

Goal 1 (for a duration of 3 sessions) was to design a bridge that spans the river and achieves the lowest mass possible (less than 175 kg) while supporting medium load at middle of each span and a factor of safety greater than 1.25. The "factor of safety" is a standard dependent characteristic of a structure studied in mechanical engineering which was automatically recomputed after every modification.

Goal 2 (for a duration of 1 session) was to design a bridge that spans river and achieves the lowest mass possible (less than 350 kg) while supporting medium load at middle of each span and a factor of safety greater than 1.25.

Goal 3 (for a duration of 2 sessions) was to design a bridge that spans river and achieves the lowest mass possible (less than 200 kg) while supporting medium load at middle of each span and has a factor of safety greater than 1.25. Additionally, participants had to ensure their bridge did not overlap the orange region (see panel 2, Figure 1).

The experiment was conducted in a cooperative setting. Participants were working on a common task (even though they each produced separate solutions), each having the same role in the team. Students constructed and tested solutions individually, they were however encouraged to discuss (in person) and share solutions (through the computer interface). Students communications were not recorded; however, the GUI allowed students to share designs and adopt the design of one of their teammates. The dataset did not include when a design was adopted but included each student designs timestamped (we used as an inference of when designs were shared). The best design of one team member at any time was used to assess the team performance. Design that met mass and factor of safety constraints were compared by weight to determine the best design. Best designs of each team were then ranked. Highest performing teams and lowest performing teams were identified as well as a group of middle teams ("other teams").

3.2 Hostage Rescue Domain

In this study[1], teams of mixed human and AI participants had to achieve computer simulated raids against automated opponents supporting building clearing efforts. Sixteen human participants were assigned to four teams (each with four human participants and additional AIs). Two parameters changed (i.e., the size of the team and the complexity of events) resulting in four different types of scenarios being administered to each team. The four different types of scenarios were: "Small and Simple", "Small and Complex", "Large and Simple" and "Large and Complex". The first parameter, with conditions "Small" and "Large", referred to the sizes of both teams and opponent forces, both of which were reduced in the "Small" condition. In the "Simple" condition, scenario events included shots fired, explosions, and deaths. In the "Complex" condition, IED and signal jamming events occurred, in addition to the "Simple" condition events.

A map of a simulation's environment is shown in Figure 2. Participants were informed that hostages were held in building K2. To release hostages, players had to occupy buildings K1 and K3 and then take K2. Southern entrances of K1, K3 were blocked to force soldiers to the northern entrance where opponent forces would ambush them in A5.

The human participants were ROTC students. Each was assigned a different role in the task: Squad Leader, Fire Team 1 Leader, Fire Team 2 Leader, and Fire Team 3 Leader. The participants' goal was to clear each target sector and eliminate opponent forces as they were encountered while

[1] DARPA A-Teams project data collected in August 2018. See Molineaux & Cox (2019) for computational details.

minimizing casualties. Failure happened when the number of human players was reduced too far to continue. The content and pattern of communications between participants were not included in the dataset.

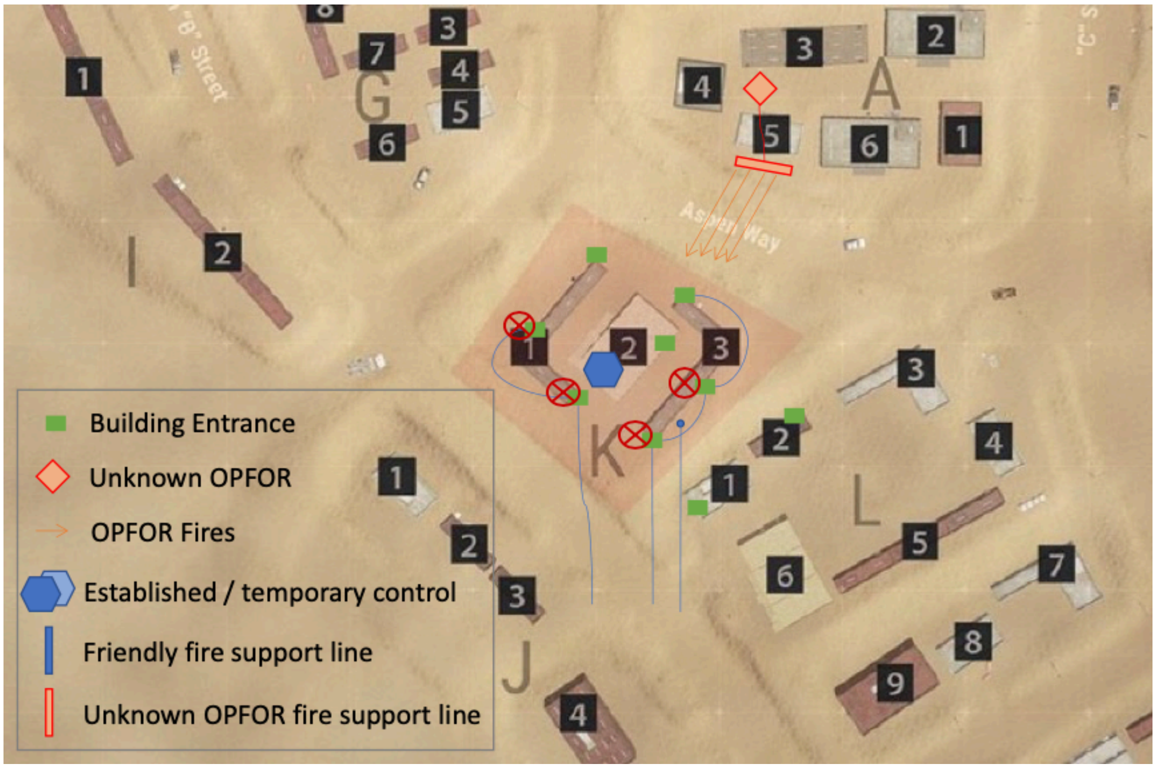

Figure 2. The hostage rescue domain: safe release of a set of hostages in building 2 within the K-Block region. Red crosses = blocked entrances; OPFOR = opposing forces.

4. Method

Following ITC, we posit that team member interaction, communications content and communication patterns are indicators of team cognition. Because, in the two studies we analyzed (bridge design and hostage rescue), direct observations of communication were not available, we posited that implicit communication traces (design in the first domain or physical dispersion and direction in the second domain), could be used as an indicator of team cognition. In each domain, we identified observables that could be used to analyze coordination. Not having access to participants' communication, we relied on indirect indicators of their behavior. Studying two distinct domains allowed us to relate observables across domains, but also to examine how the manifestation of coordination itself was different in the two domains.

In the design domain, where participants could decide to adopt the design of one of their teammates through the GUI, we inferred coordination from solution similarities. We used the intermediate and final designs that team members produced as direct indicators of team members' implicit representation of the problem. In the hostage rescue domain, the heterogeneity of the team (i.e., teammates with different roles), meant that the problem representation for each teammate might be tailored by its role and therefore vary between teammates. Teammates cooperated toward the same end goal but not necessarily with the same access to information. An indirect proxy for the team's problem representation was used: we assumed that team coordination was indicated by metrics of how the team cooperated towards the same goal, such as proximity, speed, etc.

4.1 The Bridge Design

We used the time series of designs from McComb, Cagan & Kotovsky (2018) to infer a participant's behavior. Without access to communication content and pattern, we used design similarity as an inference for team communication since participants could adopt each other's design using the GUI. To look for evidence of teaming, we studied the properties and similarities between teammates' designs by converting them into graphs and analyzing how the graphs evolved throughout the experiment. Bridge designs were converted into graphs with the Igraph R package (Csardi and Nepusz, 2006).The joints of the bridge corresponded to the vertices of the graph. Edges of the graph corresponded to the members of the bridge. An "adjustment" metric was defined as the tuning of graph connection weights. A "structure" metric was defined as the addition or deletion of an edge in the graph. Those two properties allowed us to identify strategies that distinguished the highest performing teams from the lowest performing teams. While participants had the opportunity to share their designs with members of their teams using the software, they used this function rarely. They did, however, discuss designs. The conversations were not recorded, but we were able to identify evidence of teaming from the produced graphs (graph similarities). Graph similarity is another variable we extracted by converting designs into graphs (using their structural properties). Graph similarity is a measure of isomorphism between two graphs. We used graph similarities throughout the experiment to evaluate the degree of coordination in a team.

It was important for us to connect momentary (during a short range of time before a problem statement change) and overall (the whole time series) behavior of the participants in order to infer the team state at different stages of the task. We studied variation of adjustment and coordination right before and after problem state changes as they capture the moments where goals of the participants might change as participants are finalizing a design or adapting to a new problem statement. Those moments were also critical changes in the environment, which are identified as salient point for the study of team cognition in ITC.

4.2 Hostage Rescue

Recorded information for this domain included positions of both human players and AI "pawns" (i.e., computer-generated entities) and events that took place during the mission (see Table 1).

Table 1. Recorded information for the hostage rescue domain. Clear cells = authors-defined measures; Underlined cells = Simulator outputs; BluFor = blue (friendly) forces; OpFor = opposing forces.

Behavior	Description	Indicates
Acceleration.mean	BluFor mean acceleration	Reaction time
Speed.mean	BluFor mean speed	Pace
Dispersion	Mean distance between humans (fire team, squad leaders)	Proximity
EnemyNoticed	# OpFor observed	Threat Perception
angleB2Deg.sd	σ of positional angle between t_n and t_{n+1} for each human	Coordination
MinDistanceSLToOpfor	Min distance of OpFor to Squad Leader	Threat Severity
ShootingAt	# BluFor shots fired by humans	Threat Engagement
ShootingAtAll	Total # BluFor shots fired	Threat Engagement
StanceNum.mean	Ordinal mean of human stance	Threat Perception
StanceNum.sd	Ordinal σ of human stance	Coordination

Recall that while the previous domain had fixed problem statement changes, changes in this domain (including explosions, shots fired, IED activations) occurred dynamically in response to participant activity. Based on successive pawn positions, we extracted speed and acceleration data, as well as “dispersion” and “angle variation” observables that aggregate team movements (Table 2). Dispersion was defined as the average pairwise distance between participants on the human team. Angle variation was computed as the change in angle between one participant at time t_n and the same participant at time t_{n+1} (direction of movement).

Table 2. Coordination observables.

Observable	Description
Acceleration	Mean acceleration of BluFor human players
Speed	Mean speed of BluFor human players
Dispersion	Average distance between BluFor human player pawns
Angle variation	Standard deviation of the computation of the angles (direction) formed by positions at t_n and t_{n+1} of each participant

As we did for the bridge design domain, we looked at both overall and momentary behavior. In this domain, death events were considered to indicate a critical change in the environment.

4.3 Common Measures across Domains

To draw parallels between domains, we found measures (Table 3) that were applicable to both. These included distance to self in time, distance to others, problem change, solution properties and performance. As communication between participants was not available to infer participants' mental state and problem representation, we studied the following properties of the solution: positioning for rescue domain and graph properties of the design in the bridge domain. In both cases, in accord with the ITC framework, we examined team interactions before a critical change.

Differences in time of individual participant positioning in the hostage rescue domain (speed) and individual designs at times t_n and t_{n+1} were used to infer the evolution of mental state at an individual level: abrupt changes might indicate frustration in the participant and a less deliberate problem-solving process. Similarly, differences between designs and positions (distance between teammates) give us an indication of the team state (coordinated or uncoordinated) and allow us to infer mental state or representation at both team and individual levels. We studied the influence of those features on performance at times that were presumably indicative of changes in problem representation (e.g., the death of a team member).

The types of coordination observed were very different across the two domains. In the bridge domain, participants behaved in a somewhat deliberative fashion, and in the hostage rescue domain, they were highly reactive. This difference appears to be affected both by the domains' role diversity and time pressure. In the bridge design domain, time pressure was low and participant roles were identical even though sudden problem changes happened. In contrast, the hostage rescue domain was very high pressure, and individuals' roles and short-term goals were distinct (while still contributing to an overall command goal). Overall, the hostage rescue domain called for a more reactive response from participants.

Table 3. Metrics of solution properties, team organization (1st column) and their correspondence across domains (2nd and third column).

Metric	Hostage rescue	Bridge design
Solution properties (individual) at time t_n	Angle Speed Stance	Design: Number of edges Weight
Distance to teammates (Pairwise distances)	Between pawns positions: Angle team (direction) Dispersion	Between teammate designs: number of edges separating two graphs
Distance to self in time (differences at t_n, t_{n+1})	Acceleration Stance change Angle self (direction)	Design evolution Rate of design changes
Sudden disruptions (in problem representation)	Distance between pawns	In design sizes In number of edges between designs
Problem change	Death in the team	Problem statement change

5. Results

5.1 Bridge Design

We first considered the features and performance over the whole time series (Table 4). Certain activity patterns showed up repeatedly in two top-down behavioral features – "structure" (see Figure 3) and "adjustment" (Figure 4) – that easily distinguished high and low-performing teams. While average “adjustment” per team during the whole experiment correlated significantly with final ranking (r(14)=-0.53, *p*=.04), the average “structure” modification made by a team member during the whole experiment did not correlate significantly with a team’s performance ranking (r(14)=-0.08, *p*=.8). The “adjustment” correlation confirms results from (McComb, Cagan, & Kotovsky, 2015; McComb, Cagan & Kotovsky, 2018): high- and low-performing teams applied different strategies to solve the problem. Highest performing teams tended to create a simple structure that satisfied all constraints quickly, then spend more time tuning the structure, when compared to the lowest performing teams. We believe they engaged in a complete team sensemaking process: creating a simple frame, analyzing data through that frame and using to “tune their structure”. Structure tuning occured when a participant stopped adding or deleting bridge members to his design and instead modified the properties of existing members. Structures designed by the lowest performing teams were more diverse, and these teams spent more time converging on a stable design. We believe they never succeeded in finding an initial frame. A difference between novice and experts as they tackle the task is how fast and accurately, they can identify a frame. While the students were not yet experts, some students were maybe already more competent to select the adequate frames.

According to ITC, it is important to consider team cognition in face to critical changes. We considered the momentary relationship of the different features to performance (before every problem statement change). Average team adjustments correlated significantly with ranking (r(14)=-0.67, *p*=.005), indicating that teams who did more adjustments before a problem statement change were more likely to succeed. Average team structure modifications correlated significantly with ranking (r(14)=0.64, *p*=.008), indicating that teams who did more modification to the structure of their graph (addition or deletion) before a problem statement change were more likely to fail.

Table 4. Correlations bridge design (features with performing rank). Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05

Feature	Result
Global	
Adjustment	r(14)=-0.53, *p*=.04*
Structure	r(14)=-0.08, *p*=.8
Momentary	
Adjustment	r(14)=-0.67, *p* < .001**
Structure	r(14)=0.64, *p* < .001**
Graph similarity	r(14)=-0.54, *p*=.03*

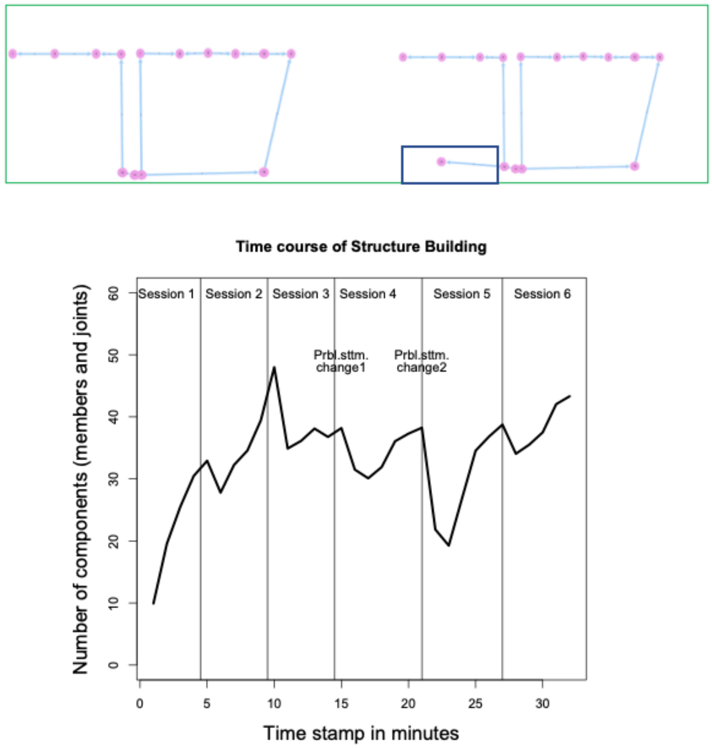

Figure 3. Team 16 participant 2 is from a lower performing team. The top panel shows two consecutive designs extracted from session 5. It illustrates the *Structure building strategy* (adding or removing components): blue rectangle indicates the added member. The bottom panel shows the evolution of the number of components (member and joints) across sessions. This participant is not settling on a design but rather adding components until the very end of session 6.

We hypothesized that graph similarity was indicative of coordination. Designs were transformed into structural graphs taking into account only Components (Members + Joints). We computed the occurrence of structurally unique graphs among participants of the same team. This showed us when similar graphs were occurring inside of the same team across participants (see Figure 5). This is close to the notion of "average pairwise similarity" which has previously been shown to be an indicator of agreement on a common solution (Wood et al., 2012). The similarity average over the entire task did not correlate significantly (see Table 5) with performance (in ranking). However, we considered this indicator in periods of time (2 minutes) before a problem statement change. The average coordination correlated significantly with the ranking of the team (r(14)=-0.54, *p*=.03) indicating that teams who were more coordinated before a problem statement change were more likely to succeed. The average coordination in high performing teams was also higher than in low performing teams (t(5.04)=-2.87, *p*=.03). Graph similarity predicted 32% of the between-team variance in performance. (Performance_rank=5.96*similarity-0.73; F(1, 14)=7.912, Adj. R sq.=0.32, *p*=.01).

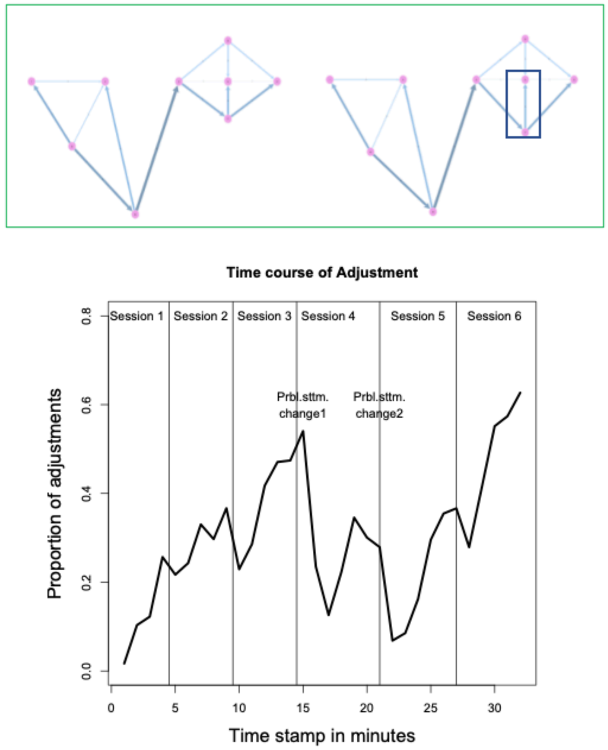

Figure 4. Team 1 participant 2 is from a higher performing team. The top panel shows two consecutive designs extracted from session 6. It illustrates the *Adjustment strategy* (changing size or position of an existing structure): blue rectangle indicates the modification. The bottom panel shows the proportion of adjustment across sessions. This participant settled early on a design in session 1 which it adjusted between session 1 and 3 and again after the problem statement change in session 5 and 6.

We then looked further at how participants explored their solution space. Were some individuals more deliberate? Did teams coordinate into being deliberative and if so, how did this affect performance? We computed the amplitude of structural change, the number of edges separating designs in the same team across time, and the number of edges separating successive graphs for the same participant in a team.

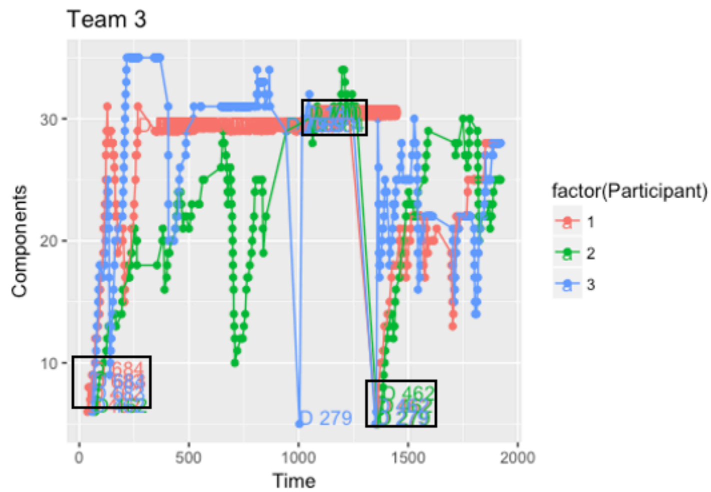

Figure 5. Team 3 is represented in this graph. Points corresponds to designs. Points with labels are designs that are structurally similar to the design of at least one other participant. Black rectangles show potential coordination moments.

Table 5. Correlation military raid simulation (features with total numbers of deaths)- Significance codes: '***' 0.001 '**' 0.01 '*' 0.05.

Feature	Result
Global	
Speed	r(14) = -0.26, *p*=.3
Dispersion	r(14) = 0.42, *p*=.1
Angle variation	r(14) = 0.09, *p*=.7
Stances variation	r(14) = -0.20, *p*=.4
Momentary	
Speed	r(14) = -0.94, *p*< .001***
Dispersion	r(14)=0.57, *p*=.02*
Angle variation	r(14)=-0.93, *p*< .001***
Stances variation	r(14)=-0.91, *p*< .001***

We first computed the pairwise difference in the number of edges separating two designs in the same team at each time. The correlation between the pairwise difference average per team and the performance of the team 2 minutes before a problem statement change was r(14)=0.6, *p*=.01. Teams who had fewer differences between their designs before the Problem Statement change had a better performance, indicating teamwork and the exploitation of the same reduced solution space. The correlation of the average of this difference per team and the performance of the team 2 minutes after a problem statement change was r(14)=0.5, *p*=.05. Participants who had fewer differences between their designs before the problem statement change had a better performance, indicating that, even as the problem was reset, teamwork continued, and a common reduced solution space was quickly adopted. The highest performing teams did not abandon their previous solution but instead tuned it to fit the new problem statement: their behavior seemed more deliberate. This might also be an ecologically rational strategy for reducing mental load by only considering a subset of a solution space, and an indication that the team sensemaking process was complete.

To measure sudden changes in an individual's exploration of the solution space, we measured the difference in the number of edges between the previous and next design individually for each participant of the team, and its standard deviation. The correlation of the average individual consecutive designs differences per team and the performance of the team 3 minutes before a problem statement change was: r(14)=0.57, *p*=.02. Individually, participants from the highest performing teams exploited a reduced solution space. The correlation of the standard deviation of individual consecutive design differences per team and the performance of the team 3 minutes before a problem statement change was: r(14)=0.71, *p*=0.002. A longer interval (3 instead of 2 minutes) was necessary to obtain enough data points for the calculation of the momentary behavior.

5.2 Hostage Rescue

We first considered the relationship of the different features to performance (as indicated by the total number of deaths) over the entire time series (see Table 5). A low total number of deaths indicates a high performance. Dispersion, angles, speed, and stances did not correlate significantly with performance. We then applied the same methodology as for the bridge design dataset by analyzing the activity in short periods before a critical event (here the death of one of the pawns in the team).

The strongest effect came from speed (see Figure 6). Slower speed had a strong correlation with the final performance (as indicated by the total number of deaths): r(14) = -0.94, *p*< .00. This effect survived even when we controlled for the size of the scenario and its complexity. It predicted 86% of in between team variances in performance (Deaths=-0.03*speed+33; F(1,14)=94.88,Adj. R sq.=0.86, *p*< .001). While dispersion correlated with deaths (r(14)=0.57, *p*=0.02), this effect disappeared as we controlled for the size of the scenario and its complexity. Angle correlated strongly with deaths (r(14)=-0.93, *p*< .001) an effect that survived as we controlled for the size of the scenario and its complexity. Variation in stances inside of a team correlated strongly negatively with deaths (r(14)=-0.91, *p*< .001), meaning that wider variations in stances correlated with a lower number of deaths. This effect also survived when we controlled for the size of the scenario and its complexity.

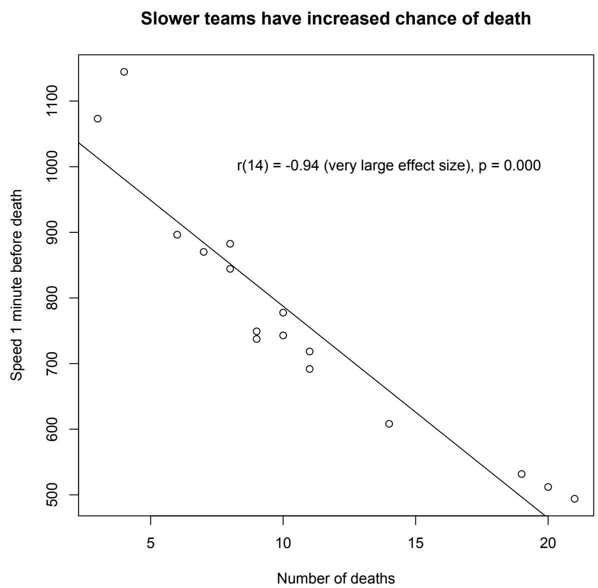

Figure 6. Correlation of Speed 1 minute before death and number of deaths.

To analyze the behavior of participants after they achieved a temporary goal, we looked at their distance to the next temporary objective (next building to be cleared) and their previous objective (previous building), averaged over all buildings captured. We specifically looked at sudden changes in distance as indicated by local maxima. A significant difference was found in the sudden changes in distance to the previous building between a team which did and did not fail $t(12.601) = 3.30$, p=.006 with teams who passed having a higher average number of disruptions. There was a difference, however not significant, in sudden changes in distance to the next building between team which passed and team which did not pass, $t(12.237) = 2.11$, p=.06 with team who pass having a higher average number of disruptions. This along with stances differences inside of the team was indicative of a more dynamic behavior in highest performing teams.

Together, dispersion, speed, stance variation, and angle predict 93% of the between-team variance in performance (Deaths= -3.445e-01 speed+1.060e+00 angle-7.289e-01 stancesd+1.238e-04 speed*Dispersion-2.995e04 Dispersion*angle+1.713e-04 Dispersion*stancesd + 2.768e+01; $F(4,11)=50.97$, Adj. R.sq.=0.93, $p < .001$).

6. Discussion

We were able to identify team strategies associated with different levels of performance in the bridge design domain and the hostage rescue domain. More specifically, in both domains we identified indirect markers of interaction and coordination which, in accord with ITC, we studied in face of critical changes.

In the bridge design domain, in agreement with the ITC framework, we observed coordination before critical moments: highest performing team showed better coordination before critical changes in the environments (problem statement changes). Additionally, we found traces of team sensemaking. Team sensemaking (Klein, Wiggins, & Dominguez, 2010) is defined as the process by which a team coordinates to explain the situation at hand. Team members which complete the sensemaking process, make decisions in a reduced decision space by framing the data they are provided and then reusing the frame they created to simplify information processing. In order to observe to which extent participants engaged in team sensemaking we identified indirect markers. Indeed, we did not have access to internalized team knowledge for either problem (e.g., explicit communication between members or individual knowledge assessments). And, we also could not rely on direct observations of communications and communication patterns since those were not available in the data. Therefore, we had to make inferences based on team behavior (indicated by design changes in the first domain and team member position changes in the second domain). We believe complete sensemaking was a determinant of performance. Indeed, highest performing teams effectively transitioned between two phases, exploration and exploitation. In the "exploration" phase they kept reframing the problem, looking for a frame that would match the given constraints, trying different bridge structures. During this phase, changes were significant and altered the main features of the solution. In "exploitation" (this is what we refer to as "tuning" in the bridge domain), participants settled and adjusted their solution. As we demonstrated, they did not do so only individually; they coordinated during "exploration" (applying a large solution space search strategy) and then coordinated in "exploitation" (adjusting the solution they had converged to). In this specific domain, the coordination measure was similarity between designs, a measure, that, as the authors of the original study (McComb, Cagan, & Kotovsky, 2015; McComb, Cagan & Kotovsky, 2018) pointed out, might not transfer to another domain .

Looking for "domain transferrable" (as opposed to "domain specific") team metrics, we attempted to explore what similar behavioral metrics would show us in two very different domains. Only three problem statement changes occurred in the span of the bridge design task. On the other hand, the rescue domain involved constant changes and the recorded behavior appeared to be more reactive than in the bridge design task. Behavioral measures reflected this key domain structure difference. We were able to observe coordination and its relationship with performance in both domains by looking at the momentary behavior before a specific critical change (i.e., death).

As in the bridge design domain, similarity in some behavioral measures indicated coordination: team members from the highest performing teams stayed close (low dispersion) to one another and moved at the same speed (see Figure 7), a behavior that might have facilitated implicit communication in the team.

Interestingly though, coordination also showed through complementarity in heterogeneous teams: in the hostage rescue domain, team members from the highest performing teams became more dissimilar in the directions and stances they were taking before critical moments (deaths). The more heterogeneous they were in stances and directions before deaths, the lower their overall number of deaths. We attribute this difference to the different roles and short-term goals participants had in the team.

In the hostage rescue domain, the best strategy across scenarios was to exhibit both similarity and complementarity: position soldiers on the battlefield close to one another and move together, while varying stance and short-term direction according to their role in the team. Coordination in this domain was also complementarity.

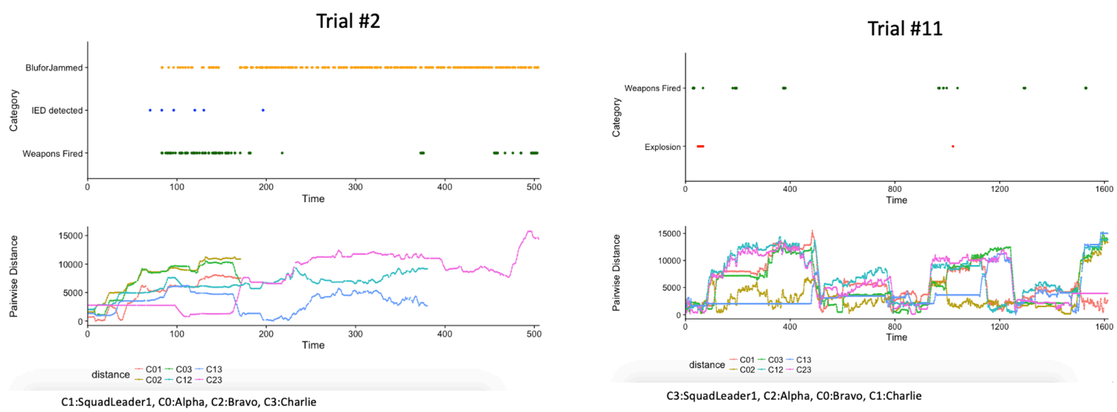

Figure 7. Pairwise distance between players(bottom) and events in a trial (top) across time for worst (Trial 2 - left) and best (Trial 11 - right) performing teams. The best performing teams show coordination in the evolution of pairwise distance: the best performing team get together after disruptions (weapons fires or explosions). Trial 11 shows this "disperse/ regroup" pattern (i.e., dispersion during a crisis and lower dispersion after the crisis).

7. Conclusion

We studied coordination across domains which are structurally different. The structure of domain and teams influenced how we measured coordination. Coordination in the bridge design task was visible through a marker of social cognition, in the similarity between designs in a team, while in the rescue domain task it was also reflected in complementarity. In the bridge design task, markers of social cognition were direct (team members design), while in the rescue domain, markers were not as easily readable. In the bridge design task, we identified different levels of coordination in the highest and the lowest performing teams. As teams were making sense of the problem, they were trying different solutions (applying different frames), and, as they figured out which frame to apply, they converged to a solution and adjusted their solution, engaging in a team sensemaking process.

In the rescue domain, we did not identify such deliberate behavior. However, we observed that in the highest performing teams, participants maintained proximity and speed (coordinating positions and movement towards their goal). Due to the different nature of the teams in this domain (different roles in the team), coordination also took a different, more complex form (high speed and close proximity while varying stance).

To capture the dynamic nature of team cognition in both domains, with a focus on Interactive Team Cognition, we studied coordination before critical changes (problem statement in the bridge design task and deaths in the rescue domain). The differences between domains allowed us to explore what observables indicate team coordination and the different types of coordination in the two domains: while coordination in the design domain was indicated by design similarities, coordination in the hostage rescue domain was indicated by how the team cooperated towards the same goal (proximity, speed, directions, etc.)

In future work, we aim to use the identified indicators of coordination to dynamically infer and predict crises in team coordination.

Acknowledgements

A shorter version of this paper was published as a poster in the non-archival online proceedings of SBP-BRIMS (Larue, et al., 2019). We thank Chris McComb for the bridge design dataset, and we thank Intific for the hostage rescue domain dataset. The material is based upon work supported by AFOSR grant #FA2386-17-1-4063, by ONR grant #N00014-18-1-2009, and by DARPA contract number N6600118C4039.

References

Cannon-Bowers, J. A., Salas, E., & Blickensderfer, E. (1999). *Toward an understanding of shared cognition*. Unpublished manuscript, Naval Air Warfare Center Training Systems Division, Orlando, Florida

Chandrasekaran, B. (1990). Design problem solving: A task analysis *AI Magazine 11*(4), 59-71.

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Journal of cognitive engineering and decision making, 3*(1), 27-46.

Cooke, N., Gorman, J. C., Myers, C. W. & Duran, J. L. (2013). Interactive team cognition. *Cognitive science 37*(2), 255-285.

Cooke, N. J. (2015). Team cognition as interaction. Current directions in psychological science, 24(6), 415-419.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, complex systems, 1695(5), 1-9.

Ghallab, M., Nau, D., & Traverso, P. (2016). *Automated Planning and Acting*. Cambridge: Cambridge University Press.

Gordon, S., Mendenhall, P., & O'Connor, B. B. (2012). *Beyond the checklist: What else health care can learn from aviation teamwork and safety*. Cornell University Press.

Klein, G., Wiggins, S., & Dominguez, C. O. (2010). Team sensemaking. *Theoretical Issues in Ergonomics Science, 11*(4), 304-320.

Larue, O., Juvina, I., Molineaux, M., Howard, B., Nichols, E., Minnery B., & Cox, M. T. (2019). Team coordination in homogeneous and heterogeneous teams. In *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS).*

Maher, M. L., Balachandran, M. B., & Zhang, D. M. (1995). *Case-based reasoning in design*. Psychology Press.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*(3), 356-376.

McComb, C., Cagan, J., & Kotovsky, K. (2015). Rolling with the punches: An examination of team performance in a design task subject to drastic changes. *Design Studies, 36*, 99-121.

McComb, C., Cagan, J., & Kotovsky, K. (2018). Data on the design of truss structures by teams of engineering students. *Data in Brief, 18*, 160-163.

Molineaux, M., & Cox, M. T. (2019). Goal-based team crisis intervention. In M. T. Cox (Ed.) *Proceedings of the Seventh Annual Conference on Advances in Cognitive Systems* (pp. 148-167). Tech. Rep. No. COLLAB 2-RT-4. Dayton, OH: Wright State University.

NTSB (1979). *United Airlines, Inc. McDonnell-Douglas DC-8-61, N8082U, Portland, Oregon, December 28,1978* (Rep. No. NTSB-AAR-79-7). Washington, DC: National Transportation Safety Board.

Salas, E. E., & Fiore, S. M. (2004). *Team cognition: Understanding the factors that drive process and performance*. American Psychological Association.

Salas, E., Fiore, S. M., & Letsky, M. P. (Eds.). (2013). *Theories of team cognition: Cross-disciplinary perspectives*. Routledge.

Santos, E., Deloach, S., & Cox, M. T. (2006). Achieving dynamic multi-commander, multi-mission planning and execution. *Applied Intelligence, 25*(3), 335-357.

Wood, M., Chen, P., Fu, K., Cagan, J., & Kotovsky, K. (2012). The role of design team interaction structure on individual and shared mental models. *Design and Cognition, 12.*