# Team coordination in homogeneous and heterogeneous teams

Othalia Larue[1], Ion Juvina[1], Matt Molineaux[2], Bruce Howard[2], Eric Nichols[2], Brad Minnery[2], Mike Cox[2]

[1] ASTECCA Laboratory, Wright State University
[2] Wright State Research Institute
Dayton, Ohio, USA
othalia.larue@wright.edu

**Abstract.** We study team interaction across two different domains: a controlled design task and a highly dynamic military raid simulation. Team interactions are measured taking into account the task dynamic. We identify cross-domain indicators of coordination and connect them to performance. In the design task showed, we observed coordinated behavior in solution space exploration and reduction. In the military raid task, we observed coordinated movement but less elaborate strategies. While coordination was observed through similarity in designs in the first domain, it was a more complex phenomenon in the second domain where complementary rather than similarity was found.

**Keywords:** Team cognition, team coordination, interaction.

## 1 Introduction

Failure and success in team interaction could lead to very different consequences in critical situations. Notably, the United Airlines 173 crash has been partially linked to failure in team communication which resulted in deficits in team cognition [1], while efficient communication and decision-making led to the "miracle on the Hudson" [2]. Effective coordination is a marker of team cognition [3,4]. Team coordination is the orchestration and timing of a sequence of interdependent actions to achieve goals and tasks [5]. The coordinated actions can vary in cognitive complexity: from moving in concert to making sense of a situation. Team sensemaking [11], notably is defined is the process by which team coordinates to explain the situation at hand.

Sharing the view of Cooke et al. [6] we see team cognition as arising in context. Team cognition emerges from individual and team factors as team members interact with their environment, the task, and each other. Meaningful team interactions occur in the presence of a critical change in the environment. A critical change is one which affects a future payoff negatively. As such, team interactions have to be measured taking into account the task dynamic.

In this paper, we present several metrics of team coordination in two very distinct task domains: the design of a truss bridge by teams of engineering students and the completion of a simulated military raid by teams of ROTC students. In the truss bridge design, 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 raid simulation, team members have different specialties and are interdependent (team members act together).

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 fit under global concepts 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 truss design, observables are related to the design properties, while in the raid domain they have to do with teammates positioning. We will consider coordination measures in the presence of a critical change in the environment (problem statement changes in the truss domain, a death in the raid simulation). Finally, we will identify what the similarities and differences between coordination properties in each of the two structurally distinct domains are.

## 2 Two task domains

We obtained data from previous studies of two different task domains. The first one, referred to as the truss design task [7,8], requires coordination between team members on an optimization problem. The other one, raid simulation [9], requires coordination on the best set of values, one for each team member.

### 2.1 Truss design

In this study [7,8], sixteen teams of three participants were assigned a truss structure (bridge spanning a chasm) design problem whose requirement are changed changes during the design process. 48 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, at the same time for each team. Changes were unexpected and required participants to adapt.

Designs were required to satisfy "factors of safety" (standard dependent characteristics of a structure studied in mechanical engineering which were automatically recomputed after every modifications) and mass constraints. 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. While students constructed and tested solutions individually, they also could discuss and share solutions through the computer interface. 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").

### 2.2 Military raid simulation

In this study [9], teams of mixed human and AI participants had to achieve computer simulated raids against automated opponents supporting buildings clearing effort. Sixteen human participants were assigned to four teams (each with four human participants and additional AIs). Two parameters changed (size of the team and 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", "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.
The human participants were ROTC students.
Each was assigned a different role in the simulation: 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 minimizing casualties. Failure happened when human pawns were reduced too far to continue.

## 3 Method

We posit that team member interaction, typically in the form of implicit communication (design or dispersion, directions), is an indicator of team cognition. In each domain, we identified observables that could be used to analyze coordination. We did not have access to participants communication, just indicators of their behavior. Studying two distinct domains allowed us to relate observables across domains, but also how the definition of coordination itself was different in the two domains. In the design domain, coordination was indicated by solution similarities. We used the intermediary and final designs that team members produced as direct indicators of team members' implicit representation of the problem. In the raid domain, 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.

### 3.1 Truss design

We used the timeseries provided [7,8] to infer participants behavior. 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. 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 member 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 (graphs similarities). Graph similarities is another variable we extracted by converting designs into graphs (using their structural properties). 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: in these moments participants are finalizing a design or adapting to a new problem statement.

### 3.2 Military raid simulation

Recorded information for this domain included positions of both humans and AI "pawns" (i.e., computer-generated entities), and events that took place during the mission. Recall, 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 1). 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 t1 and the same participant at time t2 (direction of movement).

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

**Table 1.** 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 t1 and t2 of each participant |

### 3.3 Common measures across domains

In order to draw parallels between domains, we found measures (Table 2) 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, in order to infer participants' mental state and problem representation, we studied the following properties of the solution: positioning for raid simulation, graph properties of the design in the truss domain. We examine team interactions before a critical change.

Differences in time of individual participant positioning in the military raid simulation (speed) and individual designs at time t and t+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).

**Table 2.** Correspondence between metrics across domains

| Metric | Raid simulation | Truss |
|---|---|---|
| Solution properties (individual) at time t | Angle<br>Speed<br>Stance | Design:<br>Number of edges<br>Weight |
| Distance to teammates (Pairwise distances) | Between pawns positions:<br>Angle team (direction)<br>Dispersion | Between teammate designs: number of edges separating two graphs |
| Distance to self in time (differences at t-1, t+1) | Acceleration<br>Stance change<br>Angle self (direction) | Design evolution<br>Rate of design changes |
| Sudden disruptions (in problem representation) | Distance between pawns | In design sizes<br>In number of edges between designs |
| Problem change | Death in the team | Problem statement change |

The types of coordination observed were very different across the two domains. In the truss domain, participants behaved in a somewhat deliberative fashion, and in the military raid domain, they were highly reactive. This difference appears to be affected both by the domains' role diversity and time pressure. In the truss domain, time pressure was low and participant roles were identical. In contrast, the military raid domain was very high pressure, and individuals' roles and short-term goals were distinct (while still contributing to an overall command goal).

# 4 Results

## 4.1 Truss design

We first considered the relationship of the features to performance, over the whole time series (Table 3). Certain activity patterns showed up repeatedly in the "structure" and "adjustment" features 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 team's performance ranking ($r(14)=-0.08$, $p=.8$). The "adjustment" correlation confirms results from [8] on the same dataset: 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. Structure tuning occurs when a participant stops adding or deleting member to his design as much and instead modifies 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 then 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.

We hypothesized that graph similarity was indicative of coordination. 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 [10]. The similarity average over the entire task did not correlate significantly 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.=32, $p=.01$).

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.

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 of reducing mental load by only considering a subset of a solution space. Deliberate behavior was also observed in [8] with high performing teams using increasingly more time between design change.

To measure sudden changes in an individual 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 designs 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.

### 4.2 Military raid simulation

We first considered the relationship of the different features to performance (as indicated by the total number of deaths) over the entire time series. A low total number of deaths indicates a high performance. Dispersion, angles, speed, and stances did not correlate significantly with performance (Table 4). We then applied the same methodology as for the truss 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. Slower speed had a strong correlation with the final performance (as indicated by the total number of deaths): r(14) = -0.94, $p$< .001. 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.

In order to analyze the behavior of participants after they achieved a temporary goal, we looked at their distance to the next temporary objective (next building) 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, $p< .001$).

**Table 3.** Correlations truss 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* |

**Table 4.** 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*** |

## 5 Discussion

We did not have access to internalized team knowledge for either problem. We made inferences based on team behavior (indicated by design changes in the first domain, team member position changes in the second domain).

We were able to identify team strategies associated with different levels of performance in the truss design domain. Team sensemaking [11] is defined as the process by which team coordinates to explain the situation at hand. When sensemaking completes, teams can make decisions in a reduced decision space. We believe sensemaking was successful for some teams and was a determinant of performance. 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 truss structures. During this phase, changes are significant and alter the main features of the solution. In "exploitation" (this is what we refer to as “tuning” in the truss domain) they settled and adjusted their solution. As we demonstrated, they did not do so only individually; they coordinated during “exploitation” (applying a large solution space search strategy) and then coordinated in exploitation (adjusting the solution they had converged to). Similarity between designs was the measure of coordination in this specific domain, and he authors of this study [7,8] pointed out that this might not transfer to another domain.

We attempted to explore what similar behavioral metrics would show us in a very different domain. Only three problem statement changes occurred in the span of the truss task. On the other hand, the raid simulation involved constant changes and the recorded behavior appeared to be more reactive than in the truss task. And, behavioral measures reflected this domain structure difference. We were able to observe coordination and its relationship with performance in both domains. We looked at the momentary behavior before a specific critical change (i.e., death).

As in the truss 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, a behavior that might have facilitated implicit communication in the team.

Interestingly though, coordination also showed through complementarity: 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.

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.

## 6 Conclusion

We studied coordination across domains which are structurally different. The structure of domain and teams influenced how we measured coordination. Coordination in the truss design task was visible in the similarity between designs, while in the raid simulation task it was also in complementarity. In the truss 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 figure out which frame to apply, they converged to a solution and adjusted their solution.

In the raid simulation domain, we did not identify such deliberate behavior. We, however, were able to observe that participants in the highest performing teams 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, we studied coordination before critical changes (problem statement in the truss design task and deaths in the raid simulation). 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 raid domain was indicated by how the team cooperated towards the same goal (proximity, speed, directions, etc.)

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

**Acknowledgments**. We thank McComb, C., Cagan, J., & Kotovsky, K. for the truss design dataset. We thank Intific for the military raid simulation 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

1. FAA, 1979, FAA Lessons learned, MCDONNELL-DOUGLAS, DC-8-61,
2. 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.
3. Salas, E., Fiore, S. M., & Letsky, M. P. (Eds.). (2013). Theories of team cognition: Cross-disciplinary perspectives. Routledge.
4. Salas, E. E., & Fiore, S. M. (2004). Team cognition: Understanding the factors that drive process and performance. American Psychological Association.
5. 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.
6. Cooke, Nancy J., Jamie C. Gorman, Christopher W. Myers, and Jasmine L. Duran. "Interactive team cognition." Cognitive science 37, no. 2 (2013): 255-285.
7. 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.
8. 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.
9. Intific. DARPA A-Teams directory. Full citation coming.
10. 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.
11. Klein, G., Wiggins, S., & Dominguez, C. O. (2010). Team sensemaking. Theoretical Issues in Ergonomics Science, 11(4), 304-320.