The Role of Timing of Information Front-Loading and Planning Ahead in All-Human vs. Human-Autonomy Team Performance

Matthew J. Scalia, Shiwen Zhou, David A.P. Grimm, Julie L. Harrison & Jamie C. Gorman
School of Psychology, Georgia Institute of Technology

The current study examines the effects of teams front-loading information and planning ahead through team-level communication during action phases of taskwork on team performance across all-human and human-autonomy teams (HATs) in a Remotely Piloted Aircraft System-Synthetic Task Environment (RPAS-STE). Twenty-one three-member teams (two participants teaming with either a trained experimenter or autonomous agent) flew an RPA with the goal of photographing target waypoints. Basing action phases on Information-Negotiation-Feedback (I-N-F) loops, we used the time difference between F-I as an indication of a team front-loading information. Planning ahead was hypothesized to occur in teams with longer F-I times. We found that all-human teams performed better than HATs while engaging in less front-loading. This indicates that F-I might have been measuring an aspect of team coordination related to optimal timing of action phases and flow of performing taskwork. Effective teamwork may require the right person (agent) get the right information at the right time rather than front-loading information as much as possible.

INTRODUCTION

A team is defined as two or more individuals who work interdependently and adaptively towards a shared goal (Salas et al., 1993). As technology has progressed, autonomous systems have come to be considered as teammates (McNeese et al., 2018). Teams with autonomous agents—those capable of observing an environment, acting upon an environment, and directing its activity toward a specific goal—as team members are conceptually referred to as human-autonomy teams (HATs; Chen & Barnes, 2014; McNeese et al., 2018). The perception of agents as full-fledged team members requires the assessment of HATs using the same criteria we use for all-human teams.

Traditional team process models use a series of Input-Process-Output (I-P-O) performance episodes separated by phases of taskwork (Marks et al., 2001). Of interest here are action and transition phases. In action phases, teams complete tasks that contribute directly to goal accomplishment, whereas, in transition phases, teams experience relative downtime and focus on evaluation and planning activities (e.g., preplanning; Marks et al., 2001). In addition, planning can and does occur in action phases. When teams plan while working toward task accomplishment, it is referred to as in-process planning (DeChurch & Haas, 2008; Weingart, 1992). In the current study, we examine in-process planning in the context of a Remotely Piloted Aircraft System-Synthetic Task Environment (RPAS-STE) to understand how planning contributes to performance in HATs versus all-human teams.

Like preplanning, in-process planning encompasses the iterative process of recognizing problems, gathering data, generating ideas, and evaluating and choosing a course of action (Lei et al., 2016). What distinguishes in-process planning is that it: (1) allows teams who do not have a pre-existing plan to develop a plan (Weingart, 1992); (2) allows teams with no or little task familiarity to allocate time during task performance for planning (Weingart, 1992); (3) occurs during task performance in action phases (DeChurch & Haas, 2008; Faludi, 1973; Lei et al., 2016; Weingart, 1992); and (4) provides teams the opportunity to make immediate progress on a task (Weingart, 1992). Because in-process planning reduces up-front knowledge requirements, and thereby encourages the gathering of data, we assert that information front-loading during task performance is an aspect of in-process planning (Weingart, 1992). Front-loading is the gathering of pertinent information needed to accomplish the team's goal before arriving at the goal (McNeese et al., 2018).

Team scientists argue that in-process planning occurs during a transition phase brought into existence through a period where low or no taskwork is being accomplished during an action phase (Lei et al., 2016; Marks et al., 2001). However, in highly-dynamic environments, we argue, times of low or no taskwork are not often encountered, such that waiting for a transition phase could be disastrous for team outcomes (Lei et al., 2016). Teams working in highly-dynamic work contexts, such as the RPAS-STE in the current study, inherently transition between routine and nonroutine situations without downtime, instilling the need for in-process planning to occur during action phases (Lei et al., 2016).

In dynamic contexts where teams must act immediately, team interaction patterns and processes (i.e., communication) become a viable method to measure team coordination and performance (Lei et al., 2016; Zijlstra et al., 2012; Cooke et al., 2008; Cooke et al., 2013). Taking a team cognition approach, where cognitive processing is localized 'between the heads' (BTH), team process can be measured as directly observable team interaction (Cooke et al., 2008). Indeed, a proposed model of team cognition from a BTH approach uses team communication as the medium to measure team process and directly predict team outcomes (Cooke et al., 2008). Here, we adapt this approach where front-loading via communication is used as a marker of in-process planning.

The Current Study

Teams consisting of either three human teammates or two humans and one autonomous agent teammate worked interdependently in a RPAS-STE with the goal of photographing target waypoints. In this dynamic environment, each team member assumes a role as either the pilot, navigator, or photographer and must communicate with one another to fly the RPAS and take target photos while avoiding hazardous waypoints, warnings, and alarms.

During the task, an experimenter recorded when the navigator sends information (I) about a target waypoint, when the pilot and photographer negotiate (N) airspeed and altitude for a specific target, and when the photographer sends feedback (F) that a good photo has been taken. Like Marks and colleagues' (2001) I-P-O framework, this I-N-F loop is a series of performance episodes taking place during continuous, overlapping action phases throughout the experiment. Teams enter new action phases once information about a target waypoint is sent (I). This typically occurs before the previous action phase has been completed and several targets in advance. The act of sending target information in advance of entering the target's action phase is defined as front-loading and, therefore, in-process planning. Thus, teams who engage in front-loading for targets two or more in advance are "planning ahead". However, the autonomous agent used in the experiment was incapable of entering new action phases two or more in advance of the current phase before it closed. We hypothesize that this restricts HATs to only front-load information for their current and following target making it impossible for them to "plan ahead", thereby negatively impacting team performance.

Because front-loading occurs through team-level communication (a measure of team process), we further hypothesize that team performance will be correlated with these communicative acts, as suggested by interactive team cognition (Cooke et al., 2013). Further, an experiment by DeChurch and Haas (2008) showed that in-process planning was positively related to team effectiveness (Lei et al., 2016).

Hypothesis 1: All-human teams will achieve higher performance scores than the HATs because the HATs are incapable of planning ahead (i.e., front-loading information two or more targets in advance).

Hypothesis 2: Teams with longer instances of planning ahead (i.e., more front-loading) will have higher performance scores.

METHOD

Participants

Twenty-one dyads (42 participants) were recruited from a southeastern university and surrounding areas. One team was excluded from our analyses because their data did not save properly. All teams participated in one six-hour session consisting of training and four 40-minute missions. Participants had normal or corrected-to-normal vision and were required to be fluent in English. Ages ranged from 18 to 31 years (*M* = 20.55, *SD* = 2.97) across 21 males, 20 females, and one non-binary person. Each participant was paid $10.00 per hour of participation or received course credit. The experiment was approved by the university's IRB.

Materials

The experiment was conducted in the Cognitive Engineering Research on Team Tasks-Remotely Piloted Aircraft System-Synthetic Task Environment (CERTT-RPAS-STE; Cooke & Shope, 2005). The CERTT-RPAS-STE is comprised of three task-role stations and three experimenter stations. The objective is to take photographs of color-coded strategic target waypoints while avoiding color-coded hazardous waypoints over a series of 40-minute missions.

The three task roles are (1) pilot – controls and monitors the altitude and airspeed of the Remotely Piloted Aircraft (RPA), vehicle heading, fuel, gears, and flaps, and interacts with the photographer to negotiate altitude and airspeed to take a clear picture of the target waypoints; (2) navigator – creates a dynamic flight plan and notifies the pilot of information regarding waypoints, including waypoint name, altitude restrictions, airspeed restrictions, and effective target radius; and (3) photographer – monitors and adjusts camera settings to take target photos and sends feedback to the other teammates regarding photo quality.

The pilot role was assumed by either a human experimenter or an autonomous agent ("synthetic teammate"; Ball et al., 2010). The synthetic teammate was developed using the ACT-R cognitive modeling architecture to simulate human cognition and interact with the human teammates using a text-chat interface (McNeese et al., 2018). The synthetic teammate can decide its own course of action based on its experiences during the dynamic task situation and is responsible for all aspects of the role (McNeese et al., 2018). The synthetic teammate was not developed with explicit teamwork skills. Yet, it is a critical part of the team and cannot be set aside if the team expects to perform well (McNeese et al., 2018). Participants were aware of when they were working with the synthetic teammate or the human experimenter.

The navigator and photographer roles were occupied by participants. All team members communicated using a text-chat interface. One experimenter played the role of intelligence, who communicated with the team if they asked for help, and logged information regarding communication behavior within the task.

Experimental Design

The central experimental variable—Team Type—was manipulated between-subjects at two levels: all-human and synthetic teammate. In the all-human condition, a trained human experimenter assumed the role of the pilot for all four missions. In the synthetic teammate condition, the synthetic teammate assumed the role of the pilot for the first three missions but was replaced by the human experimenter in the fourth mission. The purpose of the latter manipulation was to determine whether behaviors practiced when working with the synthetic teammate would transfer to working with a human teammate. All teams completed PowerPoint training on their roles, one hands-on training mission, and four experimental missions. Thus, Mission is defined as a within-subjects variable with four levels: Mission 1, 2, 3 and 4. There were three dependent variables—team performance, front-loading, and subjective planning ahead ratings—which are explained in the Measures section.

Procedure

Before arriving, each team was randomly assigned to an experimental condition and participants were randomized to task roles. After providing informed consent, participants completed a 30-minute interactive training PowerPoint module focusing on the participant's role, followed by a 30-minute hands-on training mission to familiarize them with the CERTT-RPAS-STE. Experimenters coached the participants while following a script to ensure each participant understood their role, the task, and how to use the text-chat interface. Teams then engaged in Missions 1-4. Short breaks were distributed in-between each Mission. Participants were then debriefed and paid for their participation.

Measures

Several measures were taken in this experiment, including team performance, team process measures, and team communication behavior.

Team performance. Team performance, an outcome-based measure, is scored out of 1,000. Teams begin with 1,000 points and points are deducted or added based on a weighted composite of team-level parameters, including time spent in alarm and warning states, fuel and memory resource consumption, number of missed targets, and rate of good target photos taken per minute.

Team process: Coordination. During the experiment, two experimenters tagged episodes of team process during action phases for each target. In the context of the experiment, an I-N-F loop (Gorman et al., 2010) is used to mark when the navigator sends information about a target waypoint (I), when the pilot and photographer negotiate airspeed and altitude for a specific target (N), and when the photographer sends feedback that a good photo has been taken (F).

Team process: Front-loading. The total time (seconds) of each I-N-F loop was calculated, where each I-N-F loop time is indicative of a team front-loading information for a specific target (McNeese et al., 2018). Specifically, the difference in time between the feedback message and the information message (F-I). We hypothesized that a team with longer F-I times engaged in more planning ahead activity.

Planning ahead subjective ratings. During the experiment, eight communication behaviors were coded by two experimenters. One of these behaviors, planning ahead, was used to mark when team members were discussing a waypoint two or more ahead of the current waypoint. We calculated the total number of planning ahead activities per mission to validate our operationalization of front-loading using the F-I metric. Inter-rater reliability for the planning ahead rating was $\kappa = .47$, qualifying as moderate agreement.

RESULTS

We ran separate 2 (Team Type) × 4 (Mission) mixed ANOVAs on team performance and front-loading. Because the within-subject data matrix resulted in missing data for one team with an outlier score on team performance and for two teams on front-loading (one with an outlier score and one that did not complete an I-N-F loop), we utilized a mixed ANOVA technique in SPSS that allowed the data to remain stacked and the remaining data from those outlier missions to remain in the analyses (Enders, n.d.). This technique analyzes the data as a multilevel model (Mission nested within Team) and uses the Satterthwaite approximation to calculate denominator degrees of freedom (UCLA: Statistical Consulting Group, n.d.). Due to the use of this technique, we were unable to generate effect sizes, aside from Cohen's *d*.

For team performance there was a significant Mission main effect, $F(3, 60.00) = 20.86$, $p < .001$, a significant Team Type main effect, $F(1, 20.00) = 19.43$, $p < .001$, and a significant Team Type × Mission interaction, $F(3, 60.00) = 4.81$, $p = .005$. For the Mission main effect, pairwise comparisons (Fisher's LSD) indicated that teams performed better in Mission 4 than any other mission ($p < .001$), but the first three missions did not significantly differ. The main effect of Team Type indicated that all-human ($M = 474.50$, $SE = 23.05$) teams performed better than the synthetic teams ($M = 330.80$, $SE = 23.05$), $p < .001$. These findings suggest performance was better when the experimenter assumed the pilot role, such that in Mission 4 both conditions had statistically equivalent performance.

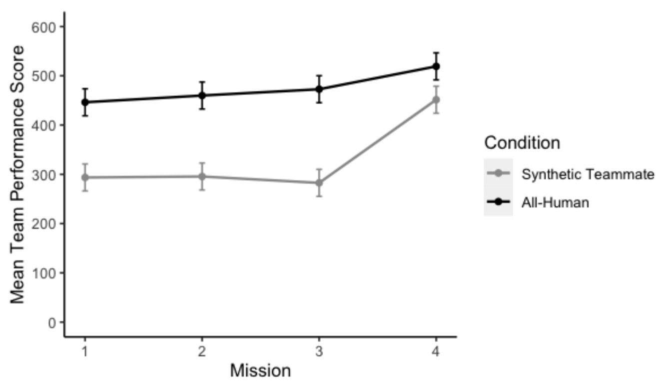

Figure 1. Mean team performance across the four missions for the two team types (bars represent standard error).

Simple effects analysis of the Team Type × Mission interaction (Figure 1), examining the effect of Mission across the different Team Types (LSD), indicated that Mission 4 outperformed all other missions for those in the synthetic Teammate condition ($p < .001$), but the first three missions did not differ. In the all-human condition, Mission 4 outperformed Mission 1 ($p = .004$) and Mission 2 ($p = .017$), however Mission 3 and 4 did not differ ($p = .60$). This suggests that the all-human teams were able to improve upon their performance from Missions 1 to 4, but the synthetic teammate may have prevented teams from improving over the first three missions. We also examined the simple effect of Team Type at different levels of Mission. The all-human teams significantly outperformed the synthetic teams for all Missions, $p < .001$, except for Mission 4, $p = .09$. This suggests that all-human teams were consistently superior and that teams that worked with the synthetic pilot were able to match the performance of the all-human teams when they partnered with the human pilot at Mission 4.

Regarding front-loading (F-I), there was a significant Mission main effect, $F(3, 58.27) = 6.97$, $p < .001$, a significant Team Type main effect, $F(1, 19.77) = 20.16$, $p < .001$, and a significant Team Type × Mission interaction, $F(3, 58.27) = 7.97$, $p < .001$. For the main effect of Mission, pairwise comparisons (Fisher's LSD) indicated that teams had variable front-loading behaviors throughout the four missions with no clear trends. Mission 1 ($M = 259.18$, $SE = 15.78$) had more front-loading than Mission 4 ($M = 206.78$, $SE = 15.41$), $p = .009$, but it did not differ from Mission 2 ($M = 247.73$, $SE = 15.78$), $p = .56$, or Mission 3 ($M = 292.55$, $SE = 15.41$), $p = .90$. Mission 2 had lower front-loading than Mission 3, $p = .02$, and Mission 4, $p = .04$. Mission 3 had more front loading than Mission 4, $p < .001$. Mission 4 had the lowest degree of front-loading, suggesting that the all-human teams exhibited less front-loading. Furthermore, front-loading, measured with the F-I metric, does not appear to be related to task experience since it had variable fluctuations.

The main effect of Team Type indicated that all-human ($M = 205.71$, $SE = 14.34$) teams engaged in less front-loading than the synthetic teammate teams ($M = 297.41$, $SE = 14.54$), $p < .001$. To further investigate this finding, we correlated the front-loading measure with the measure of planning ahead taken from team communication transcripts. Results show the two to be negatively correlated, $r(76) = -.29$, $p = .01$, which provides evidence against using F-I as a measure of planning ahead in this task. Further interpretation of the F-I measure follows in the Discussion.

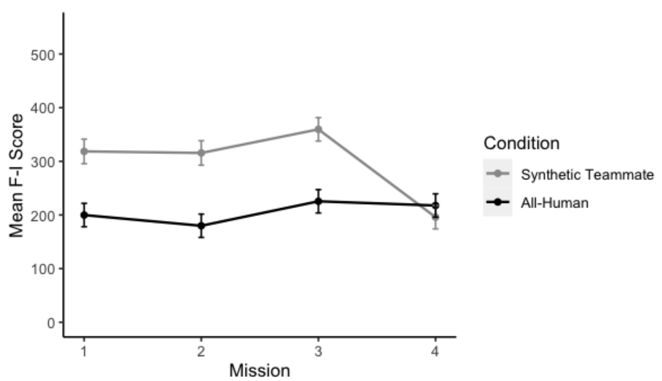

Figure 2. Mean of F-I ("front-loading") score across the four missions for the two team types (bars represent standard error).

Simple effects analysis of the Team Type × Mission interaction (Figure 2), examining the effect of Mission across the different Team Types (LSD), indicated that for synthetic teammate teams Mission 4 had a lower F-I score than all other missions, $p < .001$, but the other missions did not differ. As there were no significant differences for the all-human teams, we conclude that changes in F-I scores were primarily driven by team composition, such that teams of humans consistently had lower F-I scores. We also assessed the simple effect of Condition at the different levels of Team Type. All-human teams had significantly lower F-I scores than the synthetic teammate teams for all Missions, $p < .001$, except for Mission 4, when the synthetic teammate was replaced with the human experimenter, $p = .48$. These results are counter to our hypothesis that all-human teams would not only have higher performance but also exhibit a greater degree of information front-loading behavior than the HATs.

To further examine the efficacy of the F-I front-loading measure, we assessed its relationship with team performance and the subjective planning ahead score. We found a strong negative correlation between F-I and performance, $r(76) = -.51$, $p < .001$. Assuming that planning ahead would result in higher performance, this result suggests that larger values of F-I may not capture planning ahead, or at least capture a maladaptive and inefficient form of planning ahead. Indeed, we found that the correlation between the F-I front-loading score and the subjective planning ahead rating ($r(76) = -.29$, $p = .01$) indicating that planning ahead behaviors were associated with shorter F-I times. Moreover, there was a strong positive correlation between the subjective planning ahead ratings and team performance ($r(76) = .57$, $p < .001$). Thus, planning ahead did benefit team performance, but large F-I values were not indicative of the amount of planning ahead behaviors.

DISCUSSION

This study provides insight into the difference in team performance between all-human teams and HATs. All-human teams outperformed HATs, partially supporting Hypothesis 1. However, we found the opposite of our prediction that this difference would be due to increased ability to plan ahead by front-loading information in all-human teams as operationalized using the F-I metric. In fact, it appears that if the F-I front-loading metric does measure some aspect of planning ahead, then HATs exhibited increased levels of this aspect. However, it is likely that F-I front-loading is measuring a different aspect of team coordination related to the optimal timing and flow of performing I-N-F action loops in this experiment, such that HATs may have employed suboptimal F-I front-loading behaviors that failed to achieve these loops efficiently. Furthermore, we found that the F-I front-loading score was negatively correlated with our subjective planning ahead ratings, where the subjective planning ahead ratings were positively correlated with performance. Together, these results suggest that while differences in F-I front-loading may have accounted for performance differences between all-human teams and HATs, this was likely not related to the positive activity of planning ahead but more so to the optimal timing of action phases to effectively complete the task. In the current experiment, HATs may not have been tuned into the natural timing and flow of teamwork that achieved the best outcomes. Specifically, effective teamwork may require that the right person (agent) get the right information at the right time rather than front-loading information as much as possible.

In partial support of Hypothesis 2, F-I front-loading was correlated with performance outcomes but in a direction opposite of our prediction. In hindsight, we believe that shorter F-I times could be indicative of effective communication throughout the all-human teams. Instead of spending time processing the information provided for each target upfront and waiting to take a photo, the human operators may have been capable of integrating information for multiple targets at a time.

This may have allowed the all-human teams to process future target information more efficiently and incorporate it into the ongoing actions needed to fly the RPAS through target waypoints with increased effectiveness. It is likely that all-human teams were planning ahead, just in a more timely and efficient manner.

Alternatively, the agent's inability to fully process the I-N-F loop could have resulted in increased F-I times. Due to the limited abilities of the agent, it would occasionally and incorrectly move past the current target waypoint without negotiating with the photographer to ensure a photo was taken. As a result, the RPA could have already been headed to the next waypoint when the agent was prompted to fly the RPA back to the previously missed target to get the photograph, resulting in inflated F-I times. Specifically, when the agent then flew the RPA back towards the missed target it would cause a longer action phase to occur and, therefore, produce a larger F-I score for this specific target.

We also suspect that the shorter front-loading (F-I) times for all-human teams might be an artifact of shorter action phases for each target. That is, all-human teams processed more targets more quickly, resulting in shorter I-N-F loop times and better performance; whereas, for HATs, the longer length of front-loading (F-I) might be due to longer action phases for each target. Future research should unpack these interpretation difficulties to elucidate the effects of front-loading and planning ahead on all-human vs. HAT performance.

Finally, the pilot role for all-human teams was played by a highly-trained experimenter who was skilled in the timely pushing and pulling of target information in this task. In the future, including all-novice human teams as a control group could be an effective study design to investigate whether F-I front-loading has a range of effects involving more variation in coordination skill levels among human pilots.

Conclusion

Overall, the current study had complex tasks that required integration of information, optimal planning ahead strategies, and the ability to view the mission as a whole. The autonomous agent's coordination ability was not sufficient for reasoning about mission critical information beyond a target or two to view the mission as a whole. Rather, the autonomous agent's ability to process targets one- or two-at-a-time decreased its ability to coordinate effectively with human teammates. Importantly, this is not about natural language interaction; rather, it is about the timing of interaction.

Our current findings extend to team settings in which effective team coordination, communication, and planning ahead are critical. Autonomous agents are increasingly viewed as full-fledged team members (Fiore & Wiltshire, 2016), so understanding the limitations and potential constraints of these teaming relationships is crucial. Our findings were observed in the context of RPAS missions in a standard environment without any environmental perturbations or degraded conditions. However, there is a need to understand how teams working with autonomous agents respond in unexpected circumstances (Cooke et al., 2020). A natural next step is to extend the concepts of front-loading and planning ahead to environments susceptible to unpredictable situations and degraded conditions.

ACKNOWLEDGMENTS

The authors are thankful to Fiorella Gambetta, Yiwen Zhao, Anna Crofton, and Ansley Lee for their assistance in data collection, participant training, and data entry.

REFERENCES

Ball, J., Myers, C., Heiberg, A., Cooke, N. J., Matessa, M., Freiman, M., & Rodgers, S. (2010). The synthetic teammate project. *Computational and Mathematical Organization Theory*, *16*(3), 271–299.

Chen, J. Y., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13-29.

Cooke, N., Demir, M., McNeese, N., Gorman, J., & Myers, C. (2020). *Human–autonomy teaming in remotely piloted aircraft systems operations under degraded conditions.* Office of Naval Research Technical Report for Grant No. N00014-17-1-2382

Cooke, N. J., Gorman, J. C., & Kiekel, P. A. (2008). Communication as Team-level Cognitive Processing. In *Macrocognition in Teams* (1st ed., pp. 51–64). Routledge.

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

Cooke, N. J., & Shope, S. M. (2005). Synthetic Task Environments for Teams: CERTT's UAV-STE. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick Editor (Eds.), *Handbook of Human Factors and Ergonomics Methods* (pp. 41–46). CRC Press.

DeChurch, L. A., & Haas, C. D. (2008). Examining team planning through an episodic lens: Effects of deliberate, contingency, and reactive planning on team effectiveness. *Small Group Research, 39*(5), 542–568.

Enders, C. K. (n.d.). *Three-factor mixed ANOVA (one between- and two within-subjects factors)*. Applied Missing Data: Companion website for Applied Missing Data Analysis. http://www.appliedmissingdata.com/three-factor-mixed-anova.html

Faludi, A. (1973). *Planning theory.* New York: Pergamon Press.

Fiore, S. M., & Wiltshire, T. J. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in psychology*, *7*, 1531.

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear dynamics, psychology, and life sciences*, *14*(3), 265.

Lei, Z., Waller, M. J., Hagen, J., & Kaplan, S. (2016). Team adaptiveness in dynamic contexts: Contextualizing the roles of interaction patterns and in-process planning. *Group & Organization Management*, *41*(4), 491–525.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A Temporally Based Framework and Taxonomy of Team Processes. *The Academy of Management Review*, *26*(3), 356–376.

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors*, *60*(2), 262-273.

Salas, E., Cannon-Bowers, J. A., & Blickensderfer, E. L. (1993). Team Performance and Training Research: Emerging Principles. *Journal of the Washington Academy of Sciences*, *83*(2), 81–106.

UCLA: Statistical Consulting Group. (n.d.). *SPSS Mixed Command.* UCLA Advanced Research Computing: Statistical Methods and Data Analytics. https://stats.idre.ucla.edu/spss/seminars/spss-mixed-command/

Weingart. L. R. (1992). Impact of Group Goals, Task Component Complexity, Effort, and Planning on Group Performance. *Journal of Applied Psychology, 77*(5), 682–693.

Zijlstra, F. R. H., Waller, M. J., & Phillips, S. I. (2012). Setting the tone: Early interaction patterns in swift-starting teams as a predictor of effectiveness. *European Journal of Work and Organizational Psychology*, *21*(5), 749–777.