Communication Networks and Team Workload in a Command and Control Synthetic Task Environment

by

Craig Johnson

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2020 by the
Graduate Supervisory Committee:

Nancy Cooke, Chair
Robert Gutzwiller
Robert Gray

ARIZONA STATE UNIVERSITY

August 2020

ABSTRACT

Despite the prevalence of teams in complex sociotechnical systems, current approaches to understanding workload tend to focus on the individual operator. However, research suggests that team workload has emergent properties and is not necessarily equivalent to the aggregate of individual workload. Assessment of communications provides a means of examining aspects of team workload in highly interdependent teams. This thesis set out to explore how communications are associated with team workload and performance under high task demand in all-human and human–autonomy teams in a command and control task. A social network analysis approach was used to analyze the communications of 30 different teams, each with three members operating in a command and control task environment of over a series of five missions. Teams were assigned to conditions differentiated by their composition with either a naïve participant, a trained confederate, or a synthetic agent in the pilot role. Social network analysis measures of centralization and intensity were used to assess differences in communications between team types and under different levels of demand, and relationships between communication measures, performance, and workload distributions were also examined. Results indicated that indegree centralization was greater in the all-human control teams than in the other team types, but degree centrality standard deviation and intensity were greatest in teams with a highly trained experimenter pilot. In all three team types, the intensity of communications and degree centrality standard deviation appeared to decrease during the high demand mission, but indegree and outdegree centralization did not. Higher communication intensity was associated with more efficient target processing and more successful target photos per mission, but a clear relationship between measures of performance and decentralization of communications was not found.

ACKNOWLEDGMENTS

This research was partially supported by the Office of Naval Research (ONR), Award N000141110844. I acknowledge the large effort from the team of undergraduate and graduate researchers who conducted data collection and recruitment for this experiment, Sandia Research Corporation who developed and integrated the Remotely Piloted Aircraft System Synthetic Task Environment (RPAS-STE), and the Air Force Research Laboratory (AFRL) team who developed the synthetic agent used in this research.

TABLE OF CONTENTS

Page

LIST OF TABLES v

LIST OF FIGURES vi

LIST OF EQUATIONS vii

CHAPTER

1 INTRODUCTION 1

2 BACKGROUND 3

Workload 3

Team Workload 6

Measuring Team Workload 11

Social Network Analysis 15

The Current Study 21

3 PROJECT OVERVIEW 24

Research Questions 24

Hypotheses 24

4 METHODOLOGY 29

Participants 29

Materials and Task Environment 29

Design 33

Procedure 34

Measures 35

5 RESULTS 41

Network Summaries 41

Degree Centrality Standard Deviation 42

CHAPTER Page

Indegree Centralization 45
Outdegree Centralization 48
Intensity 50
Performance 52
Workload Distribution 60
Summary of Results 62
6 DISCUSSION 63
Summary 63
Discussion of Findings 64
Limitations 71
Conclusion 73
REFERENCES 74
APPENDIX
A RPAS-STE WORKSTATION INTERFACES 84
B PERFORMANCE SCORE CALCULATION 88
C MODIFIED NASA-TLX QUESTIONAIRE 90
D SUMMARY NETWORK DESCRIPTIONS 94
BIOGRAPHICAL SKETCH 98

LIST OF TABLES

Table Page

1. Summary of Main Variables 27
2. Summary of Hypotheses 28
3. Sequence of Events and Mission Details 35
4. Descriptive Statistics of All Observations 55
5. Overall Relationship Between Mean Intensity and Perfromance 57
6. High and Low Performing Teams in Mission 5 60
7. Centralization and Team NASA-TLX Differences Following Mission 5 61
8. Summary of Main Findings by Hypothesis 62

LIST OF FIGURES

Figure Page

1. Communication Networks Examined by Leavitt (1951) and Shaw (1964) 17
2. Diagram of the RPAS-STE 32
3. Summary of Node-level Metrics by Condition, Mission, and Position 41
4. Three-Dimensional Summary of All Observed Networks 42
5. Degree Centrality Standard Deviation Across Conditions 44
6. Degree Centrality Staandard Deviation Across Missions 44
7. Indegree Centralization Across Conditions 47
8. Indegree Centralization Across Missions 47
9. Outdegree Centralization Across Conditions 49
10. Outdegree Centralization Across Missions 49
11. Intensity Across Conditions 51
12. Intensity Across Missions 52
13. Average Target Photos Across Conditions 54
14. Target Photos Per Mission Across Missions 54
15. Correlations Matrix of All Observations 56
16. Correlation Matrix of Mission 5 with Conditions Collapsed 58

LIST OF EQUATIONS

Equation Page

1. Degree Centrality 36
2. Outdegree Centrality 37
3. Indegree Centrality 37
4. Indegree Centralization 38
5. Outdegree Centralization 38
6. Intensity 38

CHAPTER 1

INTRODUCTION

A tank crew is composed of four individuals: a tank commander, a gunner, a loader, and a driver. They all have heterogeneous roles, and unique access to information and capabilities embedded within their vehicle. When the crew needs to take collective action, the driver, gunner, tank commander, and loader all perform synchronous tasks within a short time span. However, coordinating with one another while concurrently performing individual tasks such as loading, driving, searching, and aiming can rapidly result in a state of mental overload for all of the crew members. This is especially true for the tank commander, who must integrate information from multiple sources, coordinate the collective actions of the crew, and make critical split-second decisions. When the tank commander is overloaded, it can result in delayed actions, error, and miscommunication that can easily cascade into a state of team overload. This can be catastrophic on the battlefield.

Future command and control systems are expected to increase in complexity with the advent of more autonomous and interconnected technologies (Kott & Alberts, 2017), and inclusion of artificial agents that fill roles akin to human teammates (McNeese et al., 2018). However, history has demonstrated that the integration of automation into systems can sometimes increase workload, reduce operator situation awareness, and does not always operate as expected (Bainbridge, 1983; Endsley, 2017; Sarter et al., 1997). When automation is integrated into teams, it often fails to act as a good "teammate" (Groom & Nass, 2007; Klien et al., 2004). Human-autonomy teaming presents unique challenges to many aspects of teamwork and may have impacts on workload.

The workload of individuals within these future teams will continue to be highly interdependent, and potentially more so as the timescale for planning operations decreases, and the distributed nature of operations increase (Grayson, 2018). Therefore, it is important to form a better understanding of how workload manifests in command and control teams. However, a clear conceptualization of team workload and methods for measuring it are lacking (Bowers et al., 1997; Funke et al., 2012). This thesis uses a social network analysis approach to examine the relationship between workload and team interactions in a small-scale command and control synthetic task environment. Rather than attempting to replace existing individual and team workload measures, this thesis proposes another way to examine aspects of team workload related to communication and team coordination. The goal of this work is to further our understanding of workload as a team-level construct and explore an interaction-based method for understanding how team communications are associated with different levels of task load and performance in both all-human and human–autonomy command and control teams.

CHAPTER 2

BACKGROUND

Mental workload (sometimes referred to as “cognitive workload”, or more broadly as “workload”) is a commonly evoked construct, but is difficult to define (Young et al., 2015). Nevertheless, understanding how individual workload is conceptualized lays a foundation for understanding team workload.

Workload

Mental workload can be described as a relationship between three concepts: resources, task demands, and performance (Gopher & Donchin, 1986; Hart & Staveland, 1988; Wickens et al., 2015; Wickens & Tsang, 2015; Young et al., 2015). Modern conceptions of mental workload are influenced by previous explorations into conscious and subconscious human cognitive activities (Gopher & Donchin, 1986). Information theory contributed the idea of “information" as a substrate agnostic commodity that can flow through a system and be processed, transferred, transformed and that information channels can vary in capacity in quantifiable ways (Shannon & Weaver, 1949). However, the successful experimental application of information theory to understand the limits of human cognition was limited, suggesting that more complex, structurally based theories were needed (Gopher & Donchin, 1986). These included the idea of a structural central information human processing bottleneck (Broadbent, 1958; Treisman, 1964), stages of processing, and how different signal types interacted with one another (Wickens & McCarley, 2007). Later work adopted an energy or resource metaphor to explain attention and workload, foregoing some of the difficulties in explaining and measuring workload from a structural standpoint (i.e., Kahneman, 1973; Norman & Bobrow, 1975).

Many modern theories of human mental workload consider it to be an abstract hypothetic construct without necessarily requiring the specification of detailed

mechanisms (Gopher & Donchin, 1986; Wickens, 2008). Mental resources are generally considered limited in quantity and consisting of multiple dimensions (e.g., auditory vs. visual resources, encoding vs. responding; Wickens, 2008). Automaticity, operator skill, and other individual differences affect the relationship between resource capacity and task demands (Gopher & Donchin, 1986; Kanfer & Ackerman, 1989). For example, as a task becomes more familiar, the entire task or part of it may become more automatic to an operator and require less attention to complete to the same level of performance.

Task demands are based on the interaction between the task and the operator, rather than simply the structural characteristics of the task. This can make task demands difficult to quantify (Gopher & Donchin, 1986). Nevertheless, some task characteristics related to demands include time requirements, mental and physical resource requirements, task complexity, and information processing rates (Bedwell et al., 2014; Hendy et al., 1997). In the workload context, performance is often described as a relationship between the expected and actual execution of an activity. Expected performance may be dictated by an external entity or benchmark, or it may be internally related to the goals and expectations of the individual (Young et al., 2015).

Task performance can be limited by the resources (i.e., resource-limited) or it can be limited by the data available for the task (i.e., data-limited; Norman & Bobrow, 1975). Workload as a construct is typically evoked for resource-limited task performance, however many tasks are both resource and data limited at different points. For example, a task may start as resource limited, and then be data limited once a certain level of performance is achieved (Norman & Bobrow, 1975). Multiple tasks may be completed concurrently (multitasking) or sequentially by an individual. When an operator conducts multiple tasks simultaneously (a.k.a. multitasking or timesharing), performance is related to the resource requirements of the competing tasks. Multiple Resource Theory is

an example of an attempt to define what these resources are, predict the conditions under which interference between tasks takes place, and how it might impact performance (Wickens, 2008). Sequential taskwork may take place due to the nature of the task, or when the resources are not available to complete multiple tasks concurrently. Task performance is also related to other human factors considerations such as situation awareness which are not always directly tied to workload (Wickens & Tsang, 2015).

Overload

When several processes compete for the same resources, or resource requirements increase for a single task, resource availability will eventually be surpassed. The state where task demands exceed resource availability is referred to as *overload*. The actual point or zone at which demands exceed resources is referred to as the "red line" or "red zone" (Grier et al., 2008; Wickens et al., 2015). Degradation in performance due to overload may happen in a smooth manner or as a sudden, catastrophic failure depending on the functional relationship between resources and performance (Norman & Bobrow, 1975). Residual capacity, on the other hand, refers to the extra resources which are still available when task demands are met, but overload is not occurring (Wickens et al., 2015). Residual capacity allows an operator to take on new or changing task demands, such as off-nominal events (Wickens, 2000). However, less workload, as a general principle, does not guarantee better performance. Rather workload appears to display a curvilinear relationship akin to the Yerkes-Dodson arousal curve (Broadhurst, 1957), where excessively high or low levels of workload can lead to performance loss. Excessively low levels of workload and arousal have been associated with complacency, boredom, performance decrements, and stress (Wickens & Tsang, 2015).

Team Workload

A team is traditionally defined as two or more people working together towards common goals (Salas et al., 1992). Teams are interdependent, meaning there are complementary relationships between team members that must be managed to accomplish collective tasks. This management of dependencies is referred to as coordination (Malone & Crowston, 1994). When the capacity for a task is not possessed independently by a teammate, coordination may be required or otherwise beneficial for collective performance (Johnson et al., 2014). The pooling of resources, diversity of team member knowledge, skills and abilities, and interdependent relationships is what makes it so that a team can be "more than the sum of the parts," able to accomplish more than individuals would in isolation. However, coordinating work requires individual teammates to allocate some of their resources to teamwork activities such as communicating and monitoring (MacMillan et al., 2004). This changes the nature and quantity of the resource required when compared to working in isolation (Bowers et al., 1997).

Team workload is a conceptual relationship between task demands, the resources or capacity of a team, and task performance on the team-level—similar, but not completely isomorphic to how workload is conceptualized for individuals (Bowers et al., 1997; Funke et al., 2012). Bowers, Braun, and Morgan considered team workload as a combination of demands from *taskwork, teamwork*, and *time-sharing* (i.e., balancing teamwork and taskwork; Bowers et al., 1997). Taskwork demand refers to a team's interactions with tasks, tools, machines—traditionally associated with individual task performance. Teamwork demand refers to interactions between teammates required to exchange information, developing and maintaining communication patterns, coordinating actions, maintain social order, etc. The balance between taskwork and teamwork can impose multiple demands akin to multi-tasking (time-sharing demands)

and individuals must make trade-offs when the time or resources required are less than available (Bowers et al., 1997).

Hendy, Liao, and Milgram adopted an information processing view on workload that attempted to reconcile time and intensity-based models of workload, operator states, and information processing channel capacity (1997). Workload capacity limitations, in their view, are determined based on the processing structure of channel capacity, similar to other workload models that consider both structural and resource limitations on processing (e.g. Wickens, 2008). In individuals, this would be the neural and bodily mechanisms through which information processing unfolds. Task load is affected by the amount of information to be processed (i.e. bits of uncertainty) and the strategies chosen by the human operator. Different strategies may involve a different set or configuration of information processing structures. Hendy et al., posited that humans adapt to excessive load by changing their processing strategy, modifying their processing channel configurations to either reduce the information to be processed or increase the time available. The information processing approach to cognition has been extended to the team level, in which knowledge and information processing mechanisms are considered distributed throughout the team, and teams perform some cognitive activities, such as planning, decision making, and situation assessment as a collective. This is referred to as "team cognition" (Cooke, Gorman, Winner, et al., 2007; Salas & Fiore, 2004).

Distributed cognition, a paradigm closely related with team cognition, considers sociotechnical systems as a collective cognitive system, while also emphasizing the role of cognitive artifacts, tools, and the environment (Hutchins, 1995a, 1995b). Distributed cognition is not limited to a single individual or limited to the physical boundaries of the human body but is instead defined by the functional relationship between subunits

within a system. These subunits can be other people and artifacts within the environment (e.g., tools; Hollan et al., 2000). However, in teams and other sociotechnical systems, interaction is required between the components for collective processes to emerge. The theory of interactive team cognition proposes that *interactions* define the cognitive activity that occurs at the team-level. Interactive team cognition is an activity, not a product or a property of the team, it is tied to context, and it is best understood by examining processes at the team-level (Cooke et al., 2013; Cooke, 2015).

Considering teams as a collective cognitive system, one may consider the finite, multi-dimensional cognitive resources of the team and how those are related to performance—*team workload*. These resources, however, take a different form than just the sum of individual cognitive resources. Team workload can be considered a function of the team's interactions as well as their relationship to the environment, available tools, and the task (Funke et al., 2012). A team's capacity to allocate its resources to a task depends not only on the resources available within the individual team members, tools, and environment, but also on the ability to interact and deploy those resources effectively. One of the determinants of a team's information processing capacity would therefore include who is able to interact with whom and in what manner they can interact, placing constraints on coordination (Urban et al., 1995; Young et al., 2007). In highly interdependent teams, the communication channel capacity of team interaction may be an important determinant of a team's workload capacity (the capacity of a team to do work; Bedwell et al., 2014). Coordination and allocation of resources may include modifying team activities more directly tied to workload such as adjusting plans, responsibilities, and actions to ensure that team members are working together toward the same goal (Sellers et al., 2014), or more broadly the coordination of activities directly tied to the team's tasks.

Due to taskwork dependencies and other coordination requirements, allocation of resources to team tasks can be limited by both resource *and* data limitations within the team. This contrasts somewhat with traditional views on individual workload, where data limitations are considered separate from resource limitations. Data limitations *within* the team, are essentially team resource limitations when they are a function of team interactions. For example, if a team member is waiting on direction from another (e.g., the team leader) to begin their part of the team's task, this would be viewed as a data-limited task from a traditional individual workload perspective. However, team workload differs from individual workload in that team-internal data limitations are included as an element of coordination and task allocation within the team. A team may possess the resources to accomplish some task (the task is within the team's workload capacity), but the team may be unable to properly coordinate these resources to meet task demands and find themselves in a state of team overload. This is similar to how an individual may be in a state of overload when they are unable to deploy resources to a task due to working memory capacity limitations, processing speed, or attention allocation. This ability to coordinate available resources in response to task demands is likely to be tied to team communication and coordination in highly interdependent teams.

Team Overload

Team overload is a state in which a team's resources are unable to be allocated to meet task demands, whether by an absolute lack of resources or an inability to allocate them to a task. Typically, when an individual is in an overload state, task performance suffers, or individuals may shift their strategies to compensate (Wickens et al., 2015). The operator can change the way they do the task in order to do it more efficiently, such as using decision-making heuristics, forgoing a portion of the task, or in the case of

multiple competing tasks, shedding tasks completely—ideally by priority of importance. Alternatively, they may simply allow performance decline (Wickens & Tsang, 2015). Teams are thought to show a similar relationship between overload and strategy shifts (Entin & Serfaty, 1999; Funke et al., 2012). Individual team members may exhibit backup behaviors, where they monitor the performance of a teammate and step in to assist with taskwork when they perceive that someone is overloaded (Sellers et al., 2014).

Teams may also collectively change their interaction strategies to compensate when workload is high, or in anticipation of team overload (Urban et al., 1995). This can take place as an explicit concerted effort, such as redirection by a leader, or emerge implicitly from the interaction of the individuals. For example, when task demand exceeds current capacities, a team may reduce explicit coordination (Entin & Serfaty, 1999), they may reallocate tasks between individuals (Sellers et al., 2014), they may change goals or performance standards, or teams may (intentionally or not) shed tasks completely in a manner similar to individuals. Individual overload adaptations may also interact and emerge on the team-level. For example, an individual may shed a task in overload, resulting in another teammate needing to complete the task. Or if a team member with a role in top-down coordination is overloaded, it may affect their ability to coordinate the rest of the team, and cascade into a state of team overload: an internal data limitation for the team's performance.

Research has also suggested that team workload is related to other factors such as trust and situation awareness (Parasuraman et al., 2008; Tsang & Vidulich, 2006). As a hypothetical example, miscalibrated trust within a team might lead to poor allocation of tasks between members, or increased monitoring activities. Increased workload of one member may in turn be associated with a reduction in sharing of critical information,

and reduced situation awareness, which could increase workload for other members of the team.

Analyzing behaviors at the team-level, and the interaction processes within the team, presents an avenue to better understand team workload. Interaction-based measures have been demonstrated as a pragmatic approach to understanding team cognition and team processes (Cooke & Gorman, 2009). Metrics which rely on observable behaviors such as explicit communications, may also provide a more concrete means for informing both human and artificial teammates on the state of the team in real time. However, methods for understanding team workload using observable behaviors are lacking.

Measuring Team Workload

This section describes some of the current approaches to assessing individual and team workload.

Primary and Secondary Task Performance

The first, and most straightforward method of measuring workload is through assessment of an operator's ability to perform their task to an acceptable level. The assumption is that as task performance suffers, this can serve as an indicator that workload demand exceeded resource availability or adequate resources were otherwise not allocated to the task. Another approach is to include a secondary task concurrently with the primary task. In general, as resource demands from the primary task increase, increased effort is required from the operator to maintain performance, which increases their workload. At a certain point, there is an inadequate allocation of resources to accomplish both tasks, and performance on either the primary or secondary task suffers (Wickens & Tsang, 2015).

Using primary or secondary task performance as a correlate to workload has the advantage of being relatively easy to implement. However, although workload and performance are directly linked, it is often difficult to discern if the effects on performance are truly due to workload limitations, and what the source of the workload limitations are. Furthermore, it can be difficult to select an appropriate secondary task (Young et al., 2015). Also, while the use of primary and secondary tasks is relatively well established for individuals, this is not the case for teams. It can be difficult to discern when changes in team performance are actually the result of resource limitations versus some other factor (e.g., situation awareness, mental models), and also difficult to discern the source of the workload (e.g., taskwork, teamwork, or a combination; Funke et al., 2012).

Subjective-self Report

The most popular way of assessing workload is through administering subjective self-report questionnaires. They typically consist of a series of questions regarding subjective experience related to workload. These can be either unidimensional, such as the Bedford Scale, which relies on perception of spare workload capacity (Roscoe, 1984) or multidimensional. The most common subjective rating scale is the NASA-Task Load Assessment (Hart & Staveland, 1988), a multi-dimensional scale which has been used in both individual and team research (Bowers & Jentsch, 2005). It consists of six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. More recently, self-report questionnaires have been developed to better differentiate between the dimension of team workload such as the Team Workload Questionnaire (TWLQ; Sellers et al., 2014).

Subjective self-report methods are popular partly because they are easy to administer. There is also a vast array of measurement scales available. Self-report

questionnaires are the most often used approach in team workload research (Bowers & Jentsch, 2005; Funke et al., 2012). However, these measures are intrusive if administered during a task and they rely on subjective recall which is prone to biases if administered afterward. Also, they may not capture the dynamics of rapidly shifting individual workload and workload distributions across a team in complex tasks, whereas research in complex team environments is likely to benefit from assessing workload at a rate which matches the rate of change in task demands (Bedwell et al., 2014). Furthermore, aggregating individual measures (e.g., calculating the mean; Bowers & Jentsch, 2005) to the team level may be inadequate for capturing some aspect of team workload if the tasks are highly interdependent. Aggregate measures may fail to capture emergent properties of team coordination related to workload such as delays in team information processing affecting multiple individuals or uneven workload distributions from bottlenecks in a single individual.

Physiological

Physiological measures (also referred to as physio-behavioral) focus on the assessment of physiological processes as an indicator of workload. They include a number of different measures such as echocardiography (ECG), eye tracking, electroencephalography (EEG), transcranial doppler ultrasonography (TCD), functional magnetic resonance imaging (fMRI) and have shown some success as indicators of individual workload (Funke et al., 2012; Vidulich & Tsang, 2012). For example, variability in the amount of time between heartbeats has been shown to be correlated with workload states (Hughes et al., 2019).

Physiological measures have been used extensively to assess individual workload, but more recently have been extended to study team workload. For example, research has explored synchronization and entropy of heart rate variability between team

members over time (Dias et al., 2019). Use of physiological indices of workload, and in particular cardiovascular measures such as heart rate variability, have the advantage of capturing continuous data which is correlated with changes in workload in a relatively unobtrusive manner (Hughes et al., 2019). However, they have been demonstrated to show a dissociation with other measures, such as subjective self-report (Young et al., 2015). Another potential drawback is that these measures alone may fail to adequately capture how team workload is tied to team interactions, such as how communications are related to the distribution of workload in a team, although measures of synchrony (e.g., Dias et al., 2019) show some potential for linking individual and team workload.

Workload Modeling and Prediction

Another approach to workload assessment includes workload modeling and prediction through task analysis and software tools. Workload modeling can be used early on in system design to reduce the need for costly revisions later in the design process, and also as a way of assessing operator load in existing systems. One of the simplest workload modeling approaches is timeline analysis, in which workload is considered as a time utilization ratio of the time required over the time available for conducting work (e.g., Parks & Boucek, 1988). Studies using this type of approach have found approximate time utilization zones in which workload experienced by operators rapidly increases such as 65–75% time utilization for an air traffic control task (Schmidt, 1978) and 70% for a multiple UAV control task (Cummings et al., 2010). Timeline analysis is useful in that it is simple and can be implemented unobtrusively, however, other more sophisticated approaches also include methods for the incorporating the patterns of utilization (Lee et al., 2011), task intensity (Hendy et al., 1997), and multiple resource capacities, task dependencies, and some aspects of team interactions (e.g., IMPRINT; Mitchell, 2000). However, as systems become larger and more complex,

modeling approaches to workload become more difficult and potentially less accurate as the margin of error compounds. Other researchers have suggested network-based approaches to modeling workload in complex systems with some limited success. One approach analyzed teams by using a set of networks specifying their tasks and details such as temporal and resource dependencies. These networks were combined with information and communication networks that specified what information was used for tasks, and how it flowed throughout the team. Workload was estimated using a combination of the number of tasks being executed and the number of information nodes required (Houghton et al., 2008).

Stanton et al. (2018) suggest that an assessment of workload on the system or team level akin to 'distributed mental workload' would need to consider (1) the workload of multiple agents, (2) how interactions between actors shapes each other's workload, (3) how different levels of workload dynamically shift throughout task performance and, (4) what wider systemic factors constrain or facilitate workload. This thesis uses a social network analysis approach to better understand how team communications networks are associated with aspects of team workload and performance under different levels of task demand in all-human and human–autonomy teams in a command and control task.

Social Network Analysis

Social network analysis (SNA) looks at the individual as a function of their relationship to other actors within their network (Borgatti et al., 2018; Wasserman et al., 1994). In a team, this places an emphasis on the *relationships* of the team interactions, and how they influence and are influenced by team processes and task demands (Katz et al., 2004). SNA may be especially useful in understanding team workload under conditions that create a close tie between communication and resource allocation in a team.

Social networks are made up of *nodes* (also called vertices) and *links* (also called edges). Nodes can represent different things and may have one or multiple dimensions. In this work nodes represent team members. Links can represent different types of relationships between nodes, also on one or multiple dimensions. In this work, links represent a single dimension of communications between agents. A variety of metrics can be derived from networks, including at the individual node level and at the global network level (Borgatti et al., 2018; Wasserman et al., 1994).

Centralization and Teams

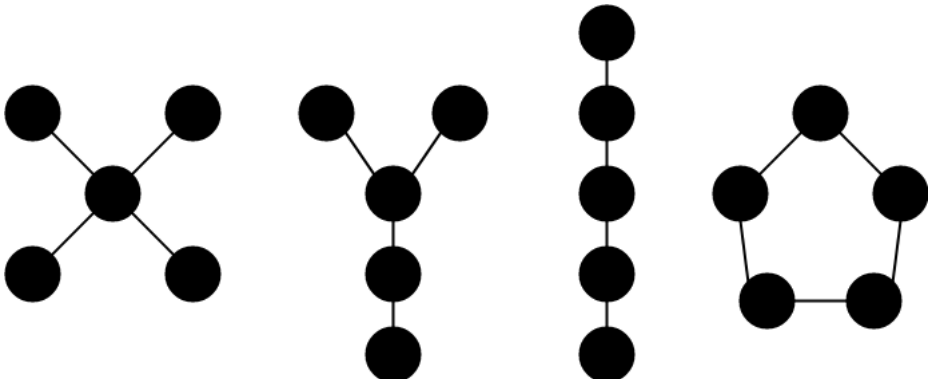
One family of measures which has been studied extensively in team and small group research is *centrality* and the network-level version *centralization* (Freeman, 1978; Katz et al., 2004). Centrality broadly describes the amount of connectedness which a node has within a network, whereas centralization is a measure of the difference in the connectedness of all the nodes in a network (Figure 1). Early work on communication networks in small groups found that imposing a structure on group communication had an influence on individual and group-level behaviors. Specifically, groups with more centralized communication structures were able to process information faster and more accurately as a team (Leavitt, 1951). However, follow-on work by Shaw et al. (1964) suggested that performance was an interaction between the communications structure and the task complexity. They reexamined previous small group experiments and found that centralized groups performed better in simple tasks, but when the task was complex, when information was distributed unevenly between members, or when the information was ambiguous, a decentralized structure performed better.

Shaw et al. hypothesized that these performance differences could be explained by what they termed *saturation* (what we would now refer to as workload): the total demands placed upon a person by communicating, manipulating data, and performing

other taskwork. They postulated that in centralized networks, the central node was more vulnerable to overload due to the greater quantity of links to other nodes and overall communication volume, and that higher complexity tasks require a greater quantity of communication and data manipulation demands. Decentralized networks, on the other hand, may require more overall communication, but are less prone to overload of individual nodes.

Later work compared teams that were structured as either in either a hierarchical or nonhierarchical manner in both high and low demand conditions. One experiment utilized the Team Performance Assessment Battery (TPAB), 5-person command and control team testbed which required participants to perform concurrent taskwork and team coordination (Bowers, Urban, & Morgan, 1992). In this experiment, it was found that the nonhierarchical teams, who were able to interact with each other in a more distributed fashion performed better under high task demand conditions in this task (Urban et al., 1995).

Figure 1
Communication Networks Examined by Leavitt (1951) and Shaw (1964).



Note. Increasing centralization from right to left.

Other work on SNA in teams focused on observing how communication networks change in response to task demands, as opposed to externally dictating network structures. One study examined task complexity and time pressure on communication networks in a group decision making task. They found that more centralized communication patterns developed when performing low-complexity tasks, and decentralized communication networks were observed when the tasks were more complex (Brown & Miller, 2000). In related observational work, surgical teams demonstrated the formation of "small-world networks" and a decentralized communication structure under periods of high task complexity (Barth et al., 2015). A small-world network is a is characterized as a balance between local and global connectedness, leading to short distances between agents, without the costs of a fully connected network (Watts & Strogatz, 1998). Small world networks have been observed as emerging in many natural and man-made networks (Barabási & Frangos, 2002). In another study, groups were assigned to either high-stress or low-stress conditions. The high-stress groups actually developed more *centralized* communication networks (Argote et al., 1989). This centralization under demand has been suggested to be linked to "threat-rigidity effects"—a maladaptive response, in which individuals tend to appeal to more centralized authority under stress, and perform in a rigid fashion (Staw et al., 1981 as cited in Katz et al., 2004).

There appear to be multiple variables that can influence the relationship between team network centralization and performance. However, it seems plausible that teams in complex environments might adopt a decentralized communication pattern when under high demands as a way to adapt their workload capacity. However, teams may not always decentralize their communications under high demand, as a more centralized structure

may be favored for simple tasks and other situations in which central nodes are less likely to be overloaded.

Communication Intensity

The amount or *intensity* of communications within a network should also be considered for understanding team workload. Intensity in this thesis refers the quantity of communication within a team over a specified amount of time. Previous work has linked increased communication intensity (and the related SNA concept of network *density*) with improved team performance (Balkundi & Harrison, 2006; Grund, 2012; Mora-Cantallops & Sicilia, 2019; Sparrowe et al., 2001). Higher team communication intensity should generally correlate with more information sharing. However, more interaction also imposes costs for the individual interlocutors, and to a lesser extent, anyone else listening over the same capacity-limited communication medium (MacMillan et al., 2004). Therefore, teams must manage communication intensity to balance the resource requirements of sending and receiving communication with the resource requirements for completing other aspects of their work.

The "voice loop" system previously used in NASA Mission Control presents an example of a system designed to manage communication network intensity in a command and control team setting. Mission Control team members were physically distributed throughout a control room where they monitored and communicated over multiple communication channels (voice loops) while also remotely managing and monitoring spacecraft systems. The Mission Control teams intentionally limited the amount of communication on some channels, such as the Flight Director Loop, to keep the network clear of excess chatter. Other channels, such as the Conference Loop could be used for in-depth and extended conversations between flight controllers and back room engineers. This system was valuable because it enabled teammates to monitor

others' activity by either listening to multiple channels in parallel or temporarily listening in to a specific channel without interrupting ongoing operations (Patterson et al., 1999). Similar systems are used in military command and control systems, where interactions may take a variety of different forms and through various different mediums such as over text and graphical interfaces (Walker et al, 2009). Limited communication channel capacity manifests in different ways on different mediums, especially in a multi-tasking environment.

The upper limits on communication channel capacities and the trade-offs which must be made when task load is high suggest that communication intensity is a vital component for understanding team workload. The upper limits on a team's communication channel capacity can be determined by many things such as the physical infrastructure, communication channel availability, and the cognitive limitations of humans and artificial systems when sending, receiving, monitoring, and transferring information within the team. Trade-offs favoring taskwork or teamwork could affect performance differently. For example, an emphasis on teamwork may increase individual workload but improve collective focus and collaboration, improving overall team performance. An emphasis on taskwork, however, may reduce individual workload, but reduce necessary information sharing and lead to poor coordination and data limitations, possibly reducing collective performance. However, similar to team centralization, the relationship between team communication intensity, team workload and performance is likely to be strongly influenced by the nature of the task and environment.

Communication Bottlenecks

The Event Analysis of System Teamwork (EAST) presents a recent example of the adaptation of social network analysis methods to human factors research (Walker et al., 2006). The approach offers a snapshot of the distributed cognition (Hutchins, 1995a) within a system by including communication, information, and task networks to describe the flow and distribution of information (Walker et al., 2006). One series of studies used the EAST methodology to examine the information sharing within submarine crews. The authors proposed that reducing information *bottlenecks* in the team communication structures may increase information sharing and reduce "production blocking"—a state in which individuals in a group are not able to produce as much output as expected because they do not have access to the requisite information/resources (Roberts et al., 2019; Stanton & Roberts, 2020). Production blocking is particularly relevant in systems in which team communication volumes are limited, such as radio networks in which only one person can speak at a time (Roberts et al., 2019), and can be conceptualized as data-limiting constraint on team workload.

Measures of network centralization and intensity may serve as a way of assessing shifts in team communications under different levels of demand. Increased centralization of a network may also suggest information bottlenecking in teams and correlate with uneven or sub-optimal workload distribution. Bottlenecks and poor workload distribution may indicate a limitation in team resource allocation, leading to high team workload. Alternatively, if task load does not exceed the capacity of a centralized network to support it, centralization may actually be indicative of an optimization strategy, in which the team focuses their communications over the "essential" channels to minimize communication overhead. Ultimately, measuring

changes in network centralization and intensity may be useful for understanding shifts in team coordination in response to task demands.

The Current Study

This study addresses team workload in a command and control context by examining the interactions of all-human and human-autonomy teams in a remotely operated aerial system synthetic task environment (RPAS-STE). The RPAS-STE was designed to elicit the cognitive and coordination aspects of an unmanned aerial system ground station but be usable by non-experts with limited training time (Cooke et al., 2004). The primary task was to coordinate as a team to safely pilot a simulated remotely piloted aircraft between a series of waypoints, while taking pictures of enemy targets under routine or roadblock conditions. Roadblocks in this task consisted of *ad hoc* targets, which were called in during the mission. To overcome a roadblock, the team was required to re-plan and take a good photo of the *ad hoc* target within a pre-specified timespan (McNeese et al., 2018). The task environment consisted of three heterogeneous and interdependent roles: navigator, photographer, and pilot. Further details on the task environment are provided in the methods section.

Workload has been examined to a limited extent in other RPAS-STE studies. Experiments on all-human teams found that the navigator and photographer roles performed differently under different levels of team load. Under high load, the navigator tended to show decrements in primary task performance but maintained stable secondary task performance (system monitoring tasks). However, the photographer role tended to maintain primary task performance with a degradation in secondary task performance under high load (Cooke, 2005). It was also found that performance was different between missions, with increases in the number of targets and amount of target restrictions resulting in teams taking fewer photos per minute despite the availability of

more targets (Cooke et al., 2007). In this task it has been found that the navigator's working memory capacity is positively correlated with team performance, and co-located navigators perceive higher workload than when they are in a separate room from the team (Cooke et al., 2007).

In another RPAS-STE study, teams received procedural, cross-training, or perturbation training prior to executing a series of missions. Teams who received the perturbation training (training intended to encourage teams to develop more adaptive coordination) performed better in the high load mission than teams who received the other forms of training (Gorman et al., 2010). This suggested that teams that had developed more adaptive (and less rigid) coordination patterns may have increased their team workload capacity.

Teaming with a synthetic agent also appear to have an impact on workload in the RPAS-STE. When participants were told they were teamed with a synthetic agent (that was actually just another participant in a separate room) they reported lower subjective workload (Cooke et al., 2016). Previous analysis of the dataset from the current experiment found that teams with a real synthetic agent filling one of the roles tended to push information to one another less overall, and instead relied on pulling information from one another. The teams with a synthetic agent also had lower team situation awareness (Demir et al., 2017), more rigid coordination patterns when compared to all-human teams (Demir et al., 2018). Other work in the RPAS-STE has focused on the related topics of team resilience and adaptation to novel and degraded conditions (Grimm et al., 2018; Lematta et al., 2019). However, team workload has not been examined using the approach presented in this thesis.

CHAPTER 3

PROJECT OVERVIEW

Research Questions

This study is motivated by two research questions: First, is a team's communication network associated with team workload and performance under high task demand in the RPAS environment? Second, how does this differ between human–autonomy and all-human teams?

Hypotheses

Hypotheses are presented in this section. Key variables and hypotheses are summarized in Table 1 and Table 2.

Hypothesis 1

All-human teams of naïve participants (control teams) and teams that include a synthetic agent (synthetic teams) will have more decentralized communications overall, whereas all-human teams with a highly trained confederate (experimenter teams) will have more centralized communications overall.

Rationale. Replacing a single teammate can have an effect on the coordination and communication of the whole team. Previous findings from this RPAS-STE experiment have shown differences in communication and coordination between naive all-human teams, teams with a highly trained experimenter following a script, and synthetic teams (McNeese et al., 2018; Demir et al., 2017; Demir et al., 2018; Grimm et al., 2018).

Control teams are expected to distribute their communications more evenly because all three members will have a similar level of skill and familiarity in the task, no clear hierarchy, and previous findings from this experiment suggest that naive all-human teams exhibit less rigid coordination behaviors (Demir et al., 2018).

Experimenter teams are expected to exhibit more centralized communication patterns due to the pilot following a set of communication rules leading to more rigid routine team coordination, the pilot exhibiting more pulling behaviors when information is not forthcoming, the potential for the participants to treat the pilot as more skilled or experienced in the task, and the actual increased skill developed by the confederate leading to the residual capacity to handle more communications.

Teams in the *synthetic* condition are expected to distribute their communications more because the limited reasoning capacity of the synthetic agent may increase the tendency for humans in that condition to coordinate and problem solve with one another rather than relying on the synthetic agent.

Hypothesis 2

All-human teams with a highly trained confederate (experimenter teams) will have the greatest communication intensity; all-human teams (control teams) will have a moderate intensity; and human–autonomy teams (synthetic teams) will have the lowest communication intensity overall.

Rationale. The experimenter teams should have an increased capacity to deal with higher communication intensity due to the confederate having expertise in the task and influencing the sharing of necessary information in a timely manner, reducing unnecessary delays in coordination. The confederate pilot's increased pulling behaviors may also increase both the pilot and the rest of the team's overall communication intensity. The synthetic team will have the least communication intensity due to the synthetic agent having limited speech comprehension and production, and its tendency to only communicate when required. Control teams will have a medium amount of communication because they do not have expertise or exceptional coordination

capacities in the task but will also have more flexibility in adapting to the needs of their teammates and the environment than the synthetic teammate.

Hypothesis 3

(a) Teams will decentralize during the high demand mission (mission 5) compared to the other missions (1–4) in order to maintain performance.

(b) Lower centralization of communications in this task will be associated with higher performance under higher demand (Mission 5).

Rationale. Previous work analyzing team communication networks has demonstrated that teams change their communications patterns in different ways based on the nature of the task demands (Argote et al., 1989; Barth et al., 2015; Brown & Miller, 2000; Katz et al., 2004; Shaw, 1964). Teams may increase their workload capacity by adapting their communication network, resulting in improved performance under high demand when compared to other teams who do not adapt to changing demands. The direction of change (centralization vs. decentralization) and how it relates to performance may be dependent on the nature of the task as well as the level of demand. Team centralization will be favored in simple tasks that can be completed more efficiently by following an optimization strategy. Decentralization will be favored for complex tasks that require heterogeneous information integration, or task demands that are high enough to overload a central node (Shaw, 1964). The increased task demand in this task is akin to an increase in information processing requirements, with some increase in task complexity in the form of replanning for *ad hoc* targets.

Hypothesis 4

(a) Increased communication intensity will be related to better performance.

(b) Reduced communication intensity will be exhibited in the high demand mission.

Rationale. Increased communication intensity reflects more information sharing, which is vital in this task. However, under high demand, explicit communications may be reduced to offset increased task load.

Hypothesis 5

Increased centralization will be associated with a larger difference in NASA-TLX scores between team members during the high demand mission.

Rationale. Centralization of communication under high demand can result in communication bottlenecks, which can result in team-internal data limitations, production blocking, and poorly distributed workload between team members.

Table 1
Summary of Main Variables.

Construct	*Measure*
Communication centralization	Degree Centrality Standard Deviation (C_{SD}) Indegree Centralization (C_I) Outdegree Centralization (C_O)
Communication intensity	Intensity
Team performance	Composite Score Target Processing Efficiency (TPE) Good Photos Per Mission
Workload distribution	Difference in NASA-TLX

Table 2
Summary of Hypotheses

Hypothesis		Expectation
1	Difference in centralization between conditions	Exp > Cont = Syn
2	Difference in intensity between conditions	Exp > Cont > Syn
3a	Difference in centralization between mission 5 and 1–4	Mission 5 < Mission 1–4
3b	Centralization related to performance in mission 5	Negative correlation between centralization and performance in mission 5
4a	Difference in intensity between mission 5 and 1–4	Mission 5 < Mission 1–4
4b	Intensity related to performance	Positive correlation between intensity and performance
5	Centralization associated with workload distribution in mission 5	Positive correlation between centralization and increased difference between teammate NASA-TLX scores

CHAPTER 4

METHODOLOGY

This thesis was conducted on data collected and reported in part elsewhere in Cooke et al. (2016), Demir et al. (2017), McNeese et al. (2018), Demir et al. (2018), and Gorman et al. (2019).

Participants

Seventy participants (60 males and 10 females, M_{age}=23.7, SD_{age}= 3.3) were recruited from Arizona State University and the surrounding areas. Participants were split into teams according to three different conditions. Thirty teams completed the experiment, with ten teams in each condition. Teams in the control condition were formed into teams of three participants, filling all three roles of navigator, pilot, and photographer. Teams in the experimenter and synthetic conditions were split into teams of two participants and were assigned to the photographer and navigator roles. In the experimenter conditions a highly trained experimenter filled the third role of pilot and in the synthetic position it was filled by a synthetic agent. Teams remained intact for the entirety of the eight-hour experiment. All participants were required to have normal or corrected-to-normal vision and be fluent in English. Each participant was compensated $10 per hour.

Materials and Task Environment

RPAS-STE Testbed

The experiment took place in the Remotely Piloted Aircraft System synthetic task environment (RPAS-STE; Cooke et al., 2004). The RPAS-STE is a team-based task environment designed to elicit some key cognitive and coordination demands of an unmanned aerial vehicle ground crew station with non-expert participants. The RPAS-STE consisted of five primary consoles. Three consoles were used as participant stations

where role-specific taskwork took place. Two other consoles were used by experimenters to monitor the experiment and code behavioral data in real-time. Each console interface had two screens, a mouse and a keyboard. Communication between participants takes place over a text-chat system on a third small screen. One of the experimenter consoles also had a chat window to communicate with the participants as needed and to inform the DEMPC of ad hoc targets at prespecified points during missions (Figure 2).

Task and Roles

The primary team task in this experiment was to maneuver a remotely piloted aircraft (RPA) as a team and take pictures of enemy targets. Constraints were embedded into the scenario including altitude and airspeed restrictions, hazardous waypoints, photo requirements, and *ad hoc* targets. Each team member had unique access to information and certain RPA functionalities, making coordination between teammates required to complete the team task. Routine coordination between the teammates followed a general sequence for the task: (1) the navigator sends waypoint information to the pilot, (2) the photographer and the pilot negotiate to achieve the proper airspeed and altitude for the current target, and (3) the photographer provides feedback to both the navigator and the pilot that they took a good picture (Figure 2; Gorman et al., 2010). Individuals were also responsible for monitoring RPA systems at their workstations such as fuel and battery levels. The roles in the experiment included navigator, pilot, and photographer.

Navigator. The navigator (Data Exploitation, Mission Planning and Communications role; DEMPC) was responsible for planning and updating the route. They did this by selecting a sequence of five waypoints at a time and sending this route to the pilot using the navigator's interface. Waypoints could be one of multiple types: (1) standard waypoints used for general navigation, (2) target waypoints, (3) restricted

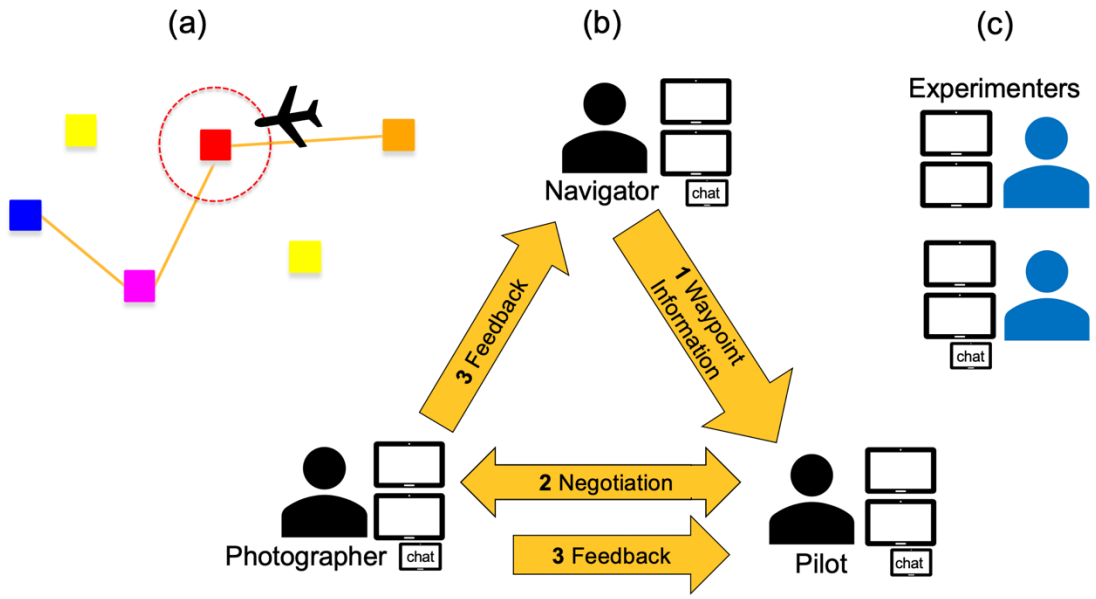
operating zone (ROZ) entry waypoints for entering ROZ boxes where targets were located, (4) ROZ exit waypoint for exiting ROZ boxes, (5) hazardous enemy waypoints, (6) environmental hazard waypoint which restricted altitude and airspeed within their radius, and (7) *ad hoc* target waypoints that were activated during the mission. Route were planned to follow the general order of (1) ROZ entry, (2) one or more target waypoints within the ROZ box, and (3) ROZ exit in order to achieve a higher score. Deviation from this order would reduce the team's score, but the navigator was not restricted from planning routes in different ways.

Photographer. The photographer (Payload Operator; PLO) was primarily responsible for taking photos of target waypoints. For each target, the photographer was provided a set of photo requirements on their console that they needed to reconcile with the RPA's state and the lighting in the environment. They did this by requesting adjustments to the altitude and airspeed and adjusting the lighting settings to match external lighting. Once all photo settings were correct and the RPA was within the radius of a target waypoint, the photographer could take a photo, inspect it for image quality and store the photo. The photographer could take multiple photos if they failed to get a good one, but this would reduce their score.

Pilot. The pilot (Air Vehicle Operator; AVO) was responsible for flying the RPA along the planned route. They did this by manipulating the RPA direction of travel, airspeed, and altitude. Coordination was required with the navigator in order to anticipate the airspeed and altitude restrictions for upcoming waypoints, and also with the photographer in order to negotiate the proper airspeed and altitude to match the photo requirements. In the experimenter condition, the pilot role was filled by a trained experimenter. The experimenter was highly trained on the task and focused on coordinating the entire team by pushing and pulling information in a timely manner

according to a script. The experimenter pilot did not make specific references to coordination strategies such as “send me the waypoint info earlier”. Rather, their influence on the team was implicit in that they pushed and pulled pertinent information in a timely manner, a form of coordination coaching (Demir et al., 2017; McNeese et al., 2018). In the synthetic condition, the pilot role was filled by an Adaptive Control of Thought-Rational (ACT-R) based computational model (synthetic agent) designed as part of a larger study (Ball et al., 2010; Myers et al., 2018). The synthetic agent was limited in speech production and comprehension.

Figure 2
Diagram of the RPAS-STE



Note. (a) A simplified representation of the virtual environment showing the aircraft entering a target radius while traveling along the planned route. (b) All three participant roles, routine interaction sequence, and workstations. (c) Two experimenters and experimenter consoles including Intel chat for injecting ad hoc targets.

Training. Training for this experiment consisted of two phases: interactive PowerPoint training, and a hands-on training mission. During the interactive training, participants watched and interacted with interface elements on a slideshow which outlined their roles and tasks. The hands-on training mission consisted of a single 40-

minute mission in which the participants practiced working together as a team to take pictures of an enemy waypoint similar to the regular mission. Experimenters followed a script to assist in clarifying tasks and roles and also answered questions during the training mission to ensure understanding before progressing to Mission 1.

Other materials. Supplementary materials were provided to assist in clarifying roles and tasks including role-specific summaries, a waypoint list for the navigator, and a notebook containing examples of good and bad photos for the photographer. Participants were also provided with quick reference guides that contained guidance on communicating with their teammates. Following this quick reference guide was especially important in the synthetic condition as it allowed the synthetic agent to understand messages better.

Design

This is focused on a specific set of variables in order to address the primary research questions. The independent variables considered for this thesis are the conditions (team composition), and the task demand. The conditions are defined by three levels: (1) the *synthetic condition* with a synthetic agent as the pilot, (2) *experimenter condition* with a highly trained experimenter as the pilot, and (3) the *control condition* with a third participant as a pilot. The task demand is split between two levels on the mission level according to the number of targets, *ad hoc* targets, restrictions, and ROZ boxes in each mission. Mission 1–4 are considered *standard demand* mission (11–13 targets, 3–4 ad hoc targets, 5 ROZ boxes), and mission five is considered a high demand mission (20 targets, 5 ad hoc targets, increased restrictions, 7 ROZ boxes; Demir et al., 2018, 2019; Grimm et al., 2018).

The dependent variables considered in this thesis include the individual and team communication flow (social network); individual workload distributions inferred from

NASA-TLX scores; and team performance measured using a composite score. The overall study design was a mixed 3 (conditions) x 5 (mission) with condition a between-subjects factor, and mission was within.

Procedure

Participants were greeted and informed consent was given. They were then randomly assigned as a team to one of the three conditions: control, experimenter, synthetic. In the control condition, participants were randomly assigned to one of three roles: pilot, navigator, or photographer. In the experimenter and synthetic conditions participants were assigned to only the navigator or photographer roles. Participants were seated at their workstations according to their roles. The photographer and navigator were located in the same room with a partition separating them. They were instructed that they could only communicate with one another over the text chat system. In the control and experimenter conditions, the pilot was located at a workstation in a separate room. In the synthetic condition the synthetic pilot was a computer program that was able to interact with the task environment. The participants were familiarized with their workstations, after which they began the interactive PowerPoint training, and then the hands-on training mission.

Following training, the participants started the first of five missions. Questionnaires were administered after the first mission to assess taskwork and teamwork knowledge and subjective workload using a modified version of the NASA-TLX (Appendix C; Hart & Staveland, 1988). Participants then completed missions 2 through 5. Missions varied in the number of (1) total targets, (2) ROZ boxes, (3) amount of waypoint constraints (route restrictions), and (4) number of *ad hoc* targets as a way of manipulating task demand (See Table 3). Mission 1 through 4 had a similar level of task demand, whereas mission five was designed to have higher task demand. Following

mission 5, workload and knowledge questionnaires were administered again, with the addition of demographics and a debrief slide. Participants were provided short breaks and offered refreshments between each mission and compensated at the end of the experiment.

Table 3

Sequence of Events and Mission Details

Event	Normal targets	*Ad hoc* targets	Total targets	ROZ Boxes	Waypoint constraints	Task demand
Training	-	-	-	-	-	-
Mission 1	8	3	11	5	Standard	Normal
Questionnaires	-	-	-	-	-	-
Mission 2	8	3	11	5	Standard	Normal
Mission 3	8	4	12	5	Standard	Normal
Mission 4	9	4	13	5	Standard	Normal
Mission 5	15	5	20	7	Increased	High
Questionnaires	-	-	-	-	-	-

Measures

A variety of measures were collected including team performance measures at the mission and target levels, team process ratings, communication flow, team verbal behaviors (McNeese et al., 2018), team situation awareness (CAST; Gorman et al., 2006), coordination metrics based on the timing of information sharing, knowledge measures based on taskwork and teamwork knowledge, and workload with NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). Only a portion of these measures are examined in this thesis, and it will be noted whenever they have been reported in previous work. Measures used in this thesis include network measures derived from communication flow, individual workload distribution measures (NASA-TLX) and team performance measures.

Social Network Analysis

Timestamped communication transcripts were automatically recorded during the experiment. Message sender, receiver, and sent time were used to derive social network analysis measures to analyze the team communication networks (Wasserman et al., 1994). Node-level measures were calculated and reported, but not analyzed. They include weighted versions of degree centrality, *indegree centrality*, and *outdegree centrality*. Due to the limitations of using SNA methods to analyze a small 3-node network, it was determined that reducing network centralization to a single measure might be inadequate. Therefore, three different measures were used to understand centralization from different perspectives: *degree centrality standard deviation* (C_{SD}), *indegree centralization* (C_I), and *outdegree centralization* (C_O). *Intensity* was used as a network-level measure of communication quantity.

Degree centrality. Individual node-level measures are first calculated in order to produce network-level measures. The first node-level measure degree centrality. The weighted version of the calculation was adopted in order to account for the number of communications, also known as node *strength* (Barrat et al., 2004; Newman, 2004). It is calculated by taking all of the incoming and outgoing links from each individual node and gives an indication of overall node activity (Freeman, 1978).

The weighted degree centrality C_w of a node i is equal to the sum of the weights of the incoming and outgoing edges as shown in Equation (1).

$$C_w(i) \sum_{i=1}^{N} w_{ji} \sum_{i=1}^{N} w_{ij} \quad (1)$$

The outdegree centrality C_{OS} or *out-strength* of a node *i* is the sum of the values of outgoing ties, which represent the number of outgoing transmissions as shown in Equation (2).

$$C_{OS}(i) \sum_{i=1}^{N} w_{ij} \tag{2}$$

The indegree centrality C_{IS} or *in-strength* C_{IS} of a node *i* is the sum of the values of incoming ties, which represent the number of communications received by a node as shown in Equation (3).

$$C_{IS}(i) \sum_{i=1}^{N} w_{ji} \tag{3}$$

Degree Centrality Standard Deviation. Centralization describes how prominent the centrality of one or a few nodes are compared to the rest of the network. A number of different measures of network centralization exist (Katz et al., 2004). However, due to the small 3-node network in this study, and the lack of measures of centralization for weighted networks on this scale, each team's standard deviation of degree centrality C_{SD} is used as one estimate of centralization. C_{SD} is calculated by taking standard deviation of the weighted degree centrality of all three team members. This gives an indication of the heterogeneity of centrality in a manner very similar to other work that uses the variance of degree centrality (Barocas et al., 2011; Snijders, 1981). High heterogeneity (high C_{SD}), means there is a large difference in the centralities of the team members, implying a more centralized communication structure, and low heterogeneity means there is less difference within the centrality of the team members, implying a less centralized communication structure.

Outdegree and Indegree Centralization. Outdegree centralization and indegree centralization and computed as shown in Equation (4) and Equation (5) respectively, where both C^{max} are the largest out-strength or in-strength observed in the network. Outdegree centralization reaches a maximum value of when all messages are received by a single node. Indegree centralization reaches its maximum value of when all messages are sent out by a single node. Both equations reach their minimum value when all messages are received and sent (respectively) in an even distribution throughout the network. Both equations are standardized by dividing by the number of nodes in the network and produce a value from 0 to 1.

$$C_O = \frac{\sum_{i=1}^{N}(C_{OS}^{max} - C_{OS}(i))}{(N-1)Intensity} \tag{4}$$

$$C_I = \frac{\sum_{i=1}^{N}(C_{IS}^{max} - C_{IS}(i))}{(N-1)Intensity} \tag{5}$$

Intensity. Density is common measure of network activity (Stanton, 2018; Wasserman et al., 1994). However, given a small network and the importance of tie weights for considering workload, traditional measures of density are not well suited for the current analysis. A global measure of network *intensity* can be considered as the average number of interactions within the team per minute of mission time, or the sum of the individual message links divided by the mission time (Grund, 2012; Mora-Cantallops & Sicilia, 2019). Considering node strengths and mission length, intensity can be computed with Equation (6).

$$Intensity = \frac{1}{\text{time}}\sum_{i=1}^{N}\left(\frac{C(i)_{IS} + C(i)_{OS}}{2}\right) = \frac{1}{time}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \tag{6}$$

Performance

Three different measures of performance are considered in this thesis in an exploratory approach: a composite score, target processing efficiency, and photos per mission.

Composite Score. The first measure of performance was a composite score designed to characterize overall team performance in this task (Cooke, Gorman, Pedersen, et al., 2007). For each mission, the performance score started at 1000 and then points were subtracted according to the weighted sum of performance penalties incurred during the mission. The performance penalties were calculated by considering time spent in alarm and warning states and missed or slow target critical and standard target photos. Timely photographs of targets and management of individual secondary tasks resulted in higher scores; therefore, this score has the benefit of considering team performance in additional to individual primary and secondary tasks (Appendix B).

Target Processing Efficiency. Target Processing Efficiency (TPE) accounts for the amount of time the aircraft time spent inside a target waypoint radius and whether or not a successful photo is attained. The score starts at 1000 and penalties are deducted as time is spent inside the target radius. A 200-point penalty is deducted for missed or bad photos. The TPE provides a target-level measure of team coordination efficiency and task performance.

Photos Per Mission. The third of measure of team performance is the number of unique good photos captured during each mission. This included both normal and *ad hoc* targets and provides a broad measure of the team's information processing rate for the primary task in the scenario. Duplicate or bad photos were not included in the measure (Cooke, Gorman, Pedersen, et al., 2007).

Workload Distribution

NASA-TLX. An adapted version of the NASA-TLX was administered to capture subjective reports of workload following two of the missions. The NASA-TLX consists of six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration (Appendix C; Hart & Staveland, 1988). It was administered to each participant following the first (baseline) mission and following the fifth (final) mission.

CHAPTER 5

RESULTS

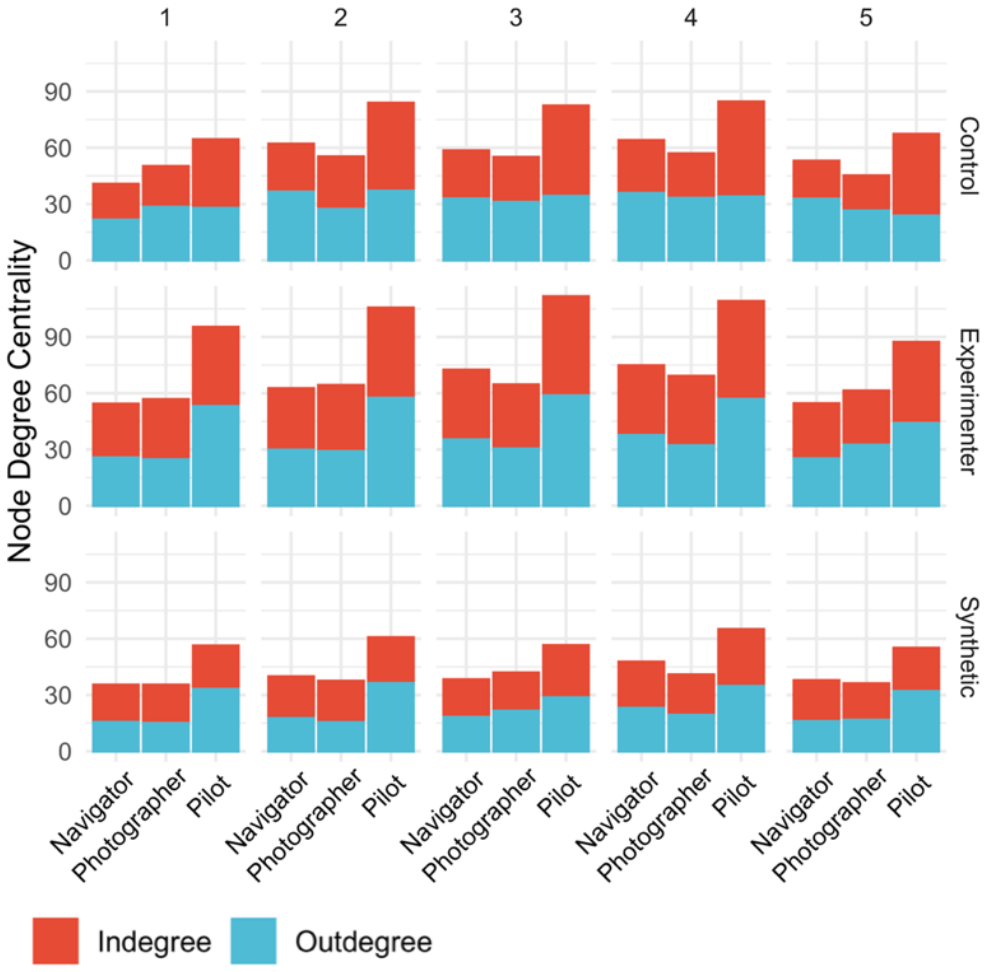
The results are presented in the following sections. Network construction, analysis, and visualization was done using the R igraph package (Csardi & Nepusz, 2006; R Core Team, 2013). Statistical analysis was conducted in SPSS and R.

Network Summaries

Node-level metrics are summarized in Figure 3.

Figure 3

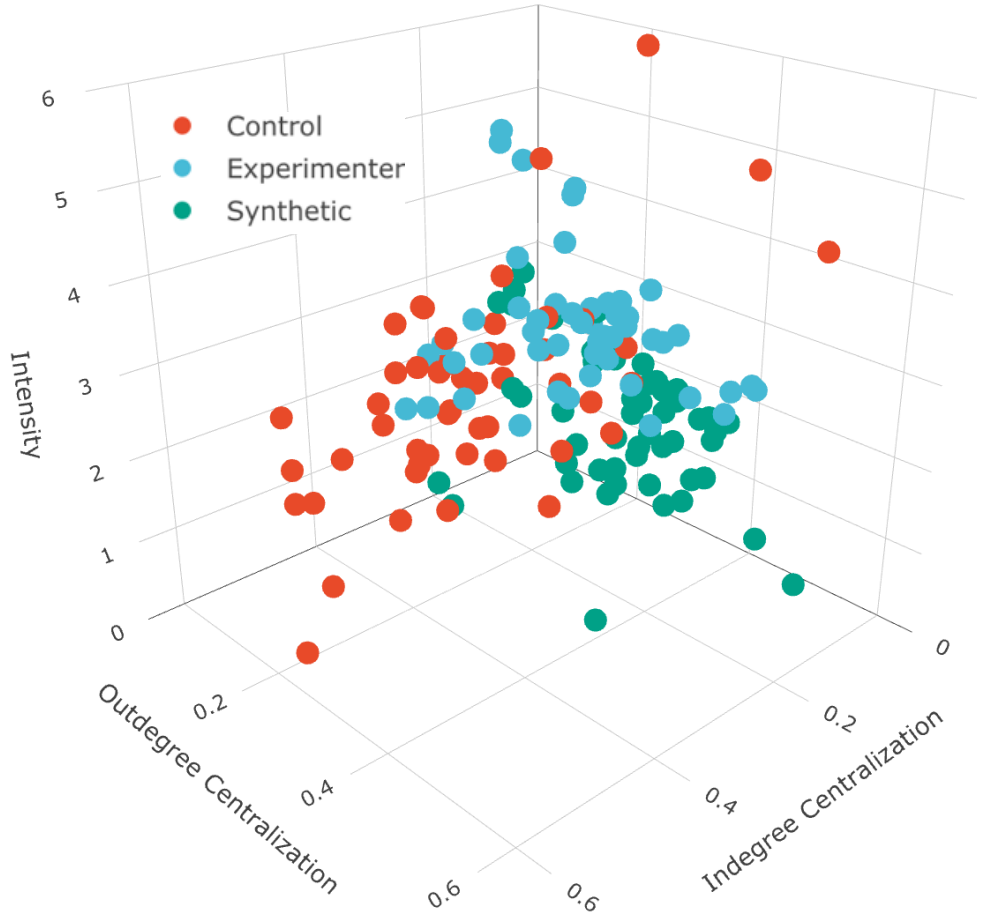
Summary of Node-level Metrics by Condition, Mission, and Position



Note. The height of each bar indicates the average weighted degree centrality (C_W) at the node-level. The top portion of each bar indicate indegree centrality (C_{IS}) and the bottom portion indicates outdegree centrality (C_{IS}). Mission numbers are shown at the top of the graph.

Network-level metrics from all 150 observations (30 teams x 5 missions) are visualized in a 3-dimensional diagram, characterizing the C_I, C_O, and intensity of each mission observed in Figure 4. Appendix D provides corresponding descriptive statistics and summary network diagrams for all missions and conditions.

Figure 4
Three-Dimensional Summary of All Observed Networks



Note. By indegree centralization (C_I), outdegree centralization (C_O), and intensity.

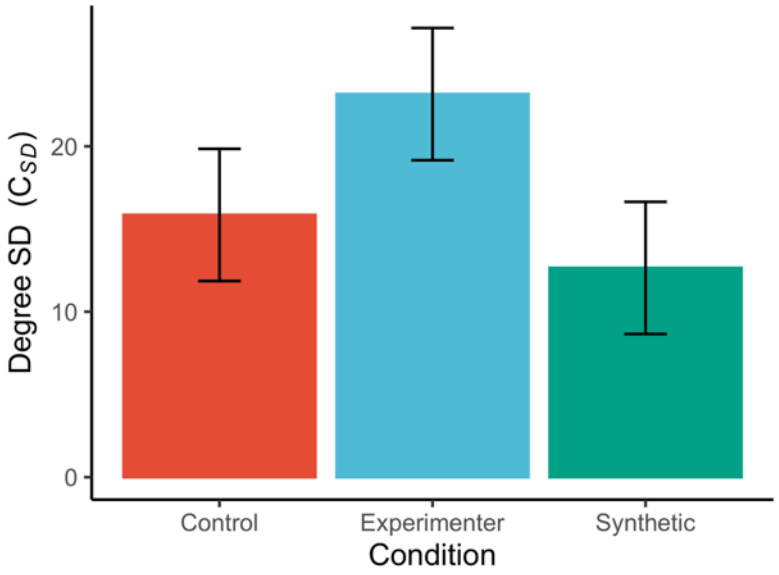
Degree Centrality Standard Deviation

In order to test for differences in degree centrality standard deviation (C_{SD}; used as a measure of centralization) between conditions and across missions, a two-way mixed ANOVA was conducted with C_{SD} as the dependent variable. There were multiple

outliers in the data, as assessed by inspection of a boxplot. Shapiro-Wilk's test indicated that C_{SD} for mission 3 and 4 in the control condition were not normally distributed ($p < .05$). However, because the outliers were within the control condition, they were less likely to be a result of experimental error, and analysis of variance is considered robust to violations of normality, the decision was made to keep the data intact and not remove outliers or transform variables. Homogeneity of variance was violated in Mission 5 ($p < .05$), however there was homogeneity of variances ($p > .05$) and covariances ($p > .05$) for all other cells, as assessed by Levene's test of homogeneity of variances and Box's M test. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 (9) = 24.542, $p = .004$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

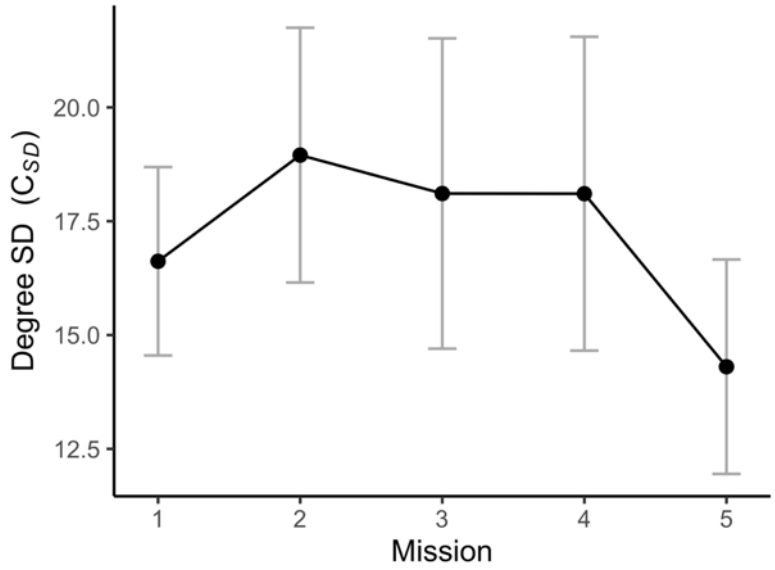
The two-way mixed ANOVA indicated there was not a statistically significant interaction between the condition and mission for mean C_{SD}, $F(5.550, 84) = 1.181$, $p = .326$, partial $\eta^2 = .080$. The main effect of condition showed a statistically significant difference in mean C_{SD}, $F(2, 27) = 7.650$, $p = .002$, partial $\eta^2 = .362$. Pairwise test results (using Least Significant difference - LSD) indicated that the experimenter teams ($M_{Exp} = 23.16$, $SD_{Exp} = 5.06$) had a higher mean C_{SD} than control teams ($M_{Cont} = 15.85$, $SD_{Cont} = 9.15$, $p = .013$) and synthetic teams ($M_{Syn} = 12.646$, $SD_{Syn} = 5.10$, $p = .001$). The mean C_{SD} for control teams was not significantly different from the synthetic teams ($p = .25$) (Figure 5). The significant main effect of condition supports the hypothesis (H1) that centralization would be different between conditions. It was predicted that the experimenter teams would have higher centralization than the control and synthetic teams, which was the case.

Figure 5
Degree Centrality Standard Deviation Across Conditions



Note: Error bars indicate 95% CI.

Figure 6
Degree Standard Deviation Across Missions



Note. Error bars indicate 95% CI.

The main effect of mission showed a statistically significant difference in mean C_{SD} across missions, $F(2.2775, 74.930) = 3.891$, $p < .014$, partial $\eta^2 = .126$. Pairwise tests results (LSD) indicated that the C_{SD} was not significantly different in Mission 2 ($M_{Mission2} = 18.95$, $SD_{Mission2} = 8.69$), Mission 3 ($M_{Mission3} = 18.11$, $SD_{Mission3} = 10.62$), and Mission 4

($M_{Mission4}$ = 18.10, $SD_{Mission4}$ = 9.72), but was significantly lower in Mission 1 ($M_{Mission1}$ = 16.62, $SD_{Mission1}$ = 7.39, p = .04) than Mission 2, and also that C_{SD} in Mission 5 ($M_{Mission5}$ = 14.30, $SD_{Mission5}$ = 6.62) was lower than Missions 2 (p = .001), mission 3 (p = .011), and Mission 4 (p = .003). There was no statistically significant difference between C_{SD} in Mission 1 and Mission 5 (p = .09) (Figure 6).

The significant main effect of the mission partially supports the hypothesis (H3a) that C_{SD} would be lower in the high task load mission than in the standard missions. Although it was not anticipated that Mission 1 would have a similar level of C_{SD} to Mission 5, one possible explanation is that team workload was higher in Mission 1 than Missions 2 through 4 because Mission 1 was the first mission immediately following the completion of training, and team skill in the task was still low.

The lack of an interaction effect between condition and mission suggests that the difference in C_{SD} between the standard missions and the higher demand mission are similar across team compositions, and that the differences in performance (reported below) between different team compositions under high demand include other factors that are not captured by C_{SD}.

Indegree Centralization

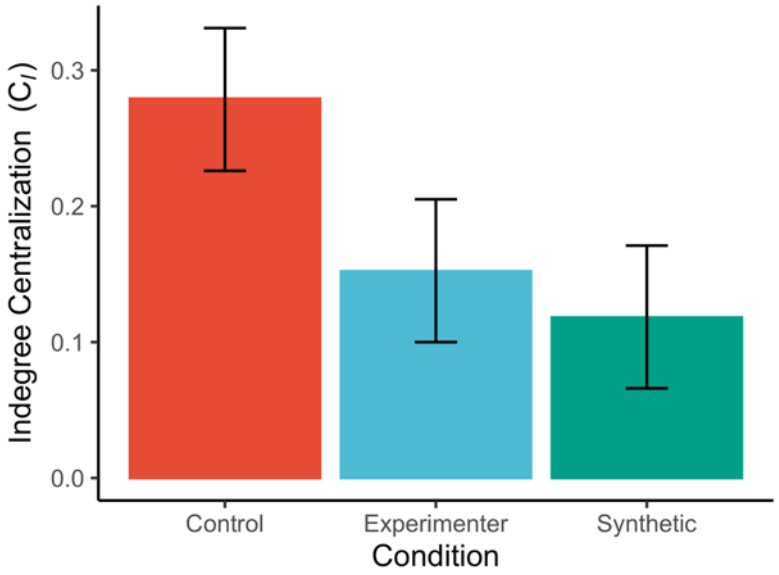
In order to a test for differences between indegree centralization (C_I) between conditions and across missions a two-way mixed ANOVA was conducted with C_I as the dependent variable. There were multiple outliers in the data, as assessed by inspection of a boxplot. Normality was violated in Mission 3 and Mission 4 for the control condition, as assessed by Shapiro-Wilk's test (p < .05). Normality was violated in Mission 1 in the experimenter and synthetic conditions, and also in Mission 5 for the experimenter condition, as assessed by Shapiro-Wilk's test (p < .05). The decision was made to keep the data intact and not remove outliers or attempt transformation of variables. Levene's

test of equality of error variances revealed that homogeneity of variance was violated for scores in Mission 4 and Mission 5 ($p < .05$). Homogeneity of covariances was violated, as assessed by Box's test of equality of covariance matrices ($p < .05$). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(9) = 19.319$, $p = .023$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

The two-way mixed ANOVA indicated that there was no significant interaction between the condition and mission for C_{IS}, $F(6.286, 84.857) = 1.603$, $p = .153$, partial $\eta^2 = .106$. The main effect of condition showed a statistically significant difference in mean C_I, $F(2,27) = 10.896$, $p < .001$, partial $\eta^2 = .447$. Pairwise tests (LSD) indicated that the control teams ($M_{Cont} = .279$, $SD_{Cont} = .142$) had a higher mean indegree centralization than experimenter teams ($M_{Exp} = .152$, $SD_{Exp} = .075$, $p = .002$) and synthetic teams ($M_{Syn} = .118$, $SD_{Syn} = .075$, $p < .001$). The mean indegree centralization for experimenter teams was not significantly different from the synthetic teams ($p = .353$; Figure 7). These findings indicate that teams in the control condition tended to have a higher number of messages received by a single individual in the team, rather than spread evenly throughout the team when compared to the synthetic and experimental conditions. Conversely, the synthetic and experimenter teams appeared to have a similar amount of indegree centralization, indicating that a smaller proportion of the total messages were received by a single individual. This is interesting considering that the team compositions, performance, and coordination in the experimenter and synthetic condition were quite different (McNeese et al., 2018). These findings are also somewhat contradictory to analysis of C_{SD}, as the experimenter teams were found to have a higher centralization when assessed using C_{SD}, but control teams had higher centralization

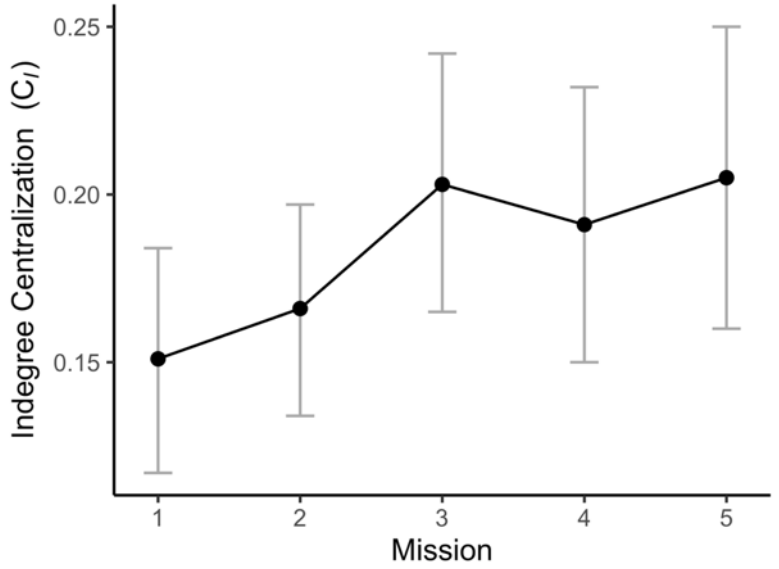
when it was measured using C_I. This suggests that these centralization measures capture different types of team communication tendencies.

Figure 7
Indegree Centralization Across Conditions



Note. Error bars indicate 95% CI.

Figure 8
Indegree Centralization Across Missions



Note. Error bars indicate 95% CI.

The main effect of mission showed a statistically significant difference in mean C_I, $F(3.143, 84.857) = 3.550$, $p = .016$, partial $\eta^2 = .116$. Pairwise tests (LSD) indicated that the C_I was significantly lower in Mission 1 ($M_{Mission1} = .151$, $SD_{Mission1} = .104$) and Mission 2 ($M_{Mission2} = .166$, $SD_{Mission2} = .095$), than in Mission 3 ($M_{Mission3} = .203$, $SD_{Mission3} = .117$), and Mission 4 ($M_{Mission4} = .191$, $SD_{Mission4} = .133$), and Mission 5 ($M_{Mission5} = .205$, $SD_{Mission5} = .156$) There was no statistically significant difference in mean C_I between Mission 1 and Mission 2 ($p = .229$), or between Mission 3, Mission 4, and Mission 5 ($p > .05$) (Figure 8). These results do not support the hypothesis that centralization would decrease in the high task demand, but instead shows an overall increase in centralization over time. However, C_{IS} only considers one part of centralization (incoming messages).

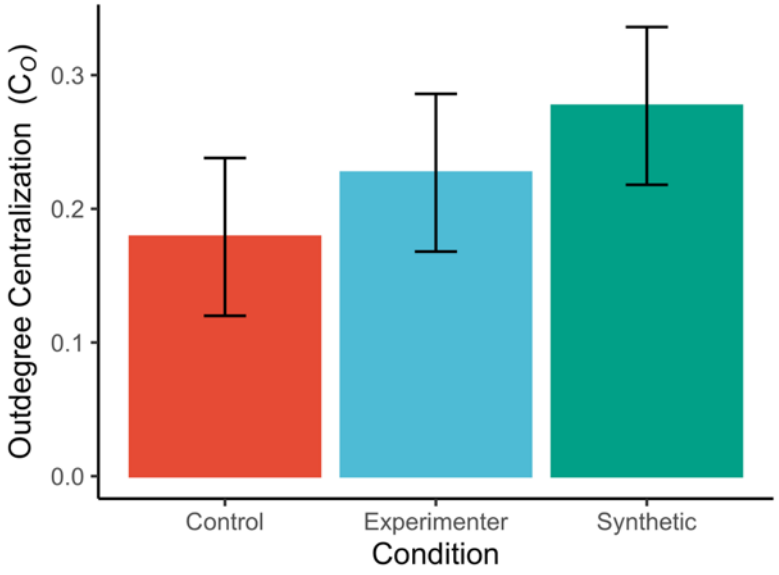
Outdegree Centralization

In order to a test for differences between outdegree centralization (C_O) between conditions and across missions a two-way mixed ANOVA was conducted with C_O as the dependent variable. There multiple outliers in the data, as assessed by inspection of a boxplot. Normality was violated in Mission 3 and Mission 4 for the control condition, as assessed by Shapiro-Wilk's test ($p < .05$). The decision was made to keep the data intact and not remove outliers or attempt transformation of variables. There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices ($p = .188$). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(9) = 21.512$, $p = .011$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

The two-way mixed ANOVA indicated that there not a statistically significant interaction between the condition and mission on C_O, $F(5.834, 78.759) = 1.437$, $p = .213$, partial $\eta^2 = .096$. The main effect of condition showed that there was not a statistically significant difference in mean C_O between the conditions, $F(5.834, 78.759) = 2.91$, $p =$

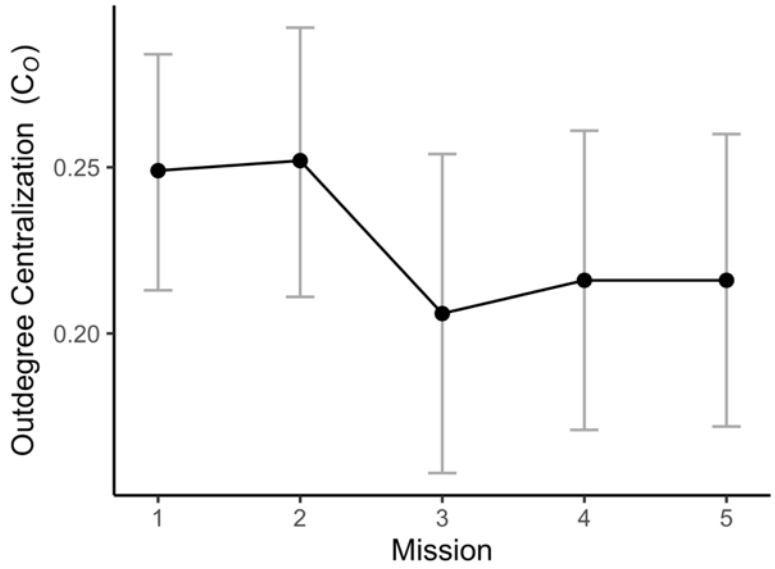
.072, partial η^2 = .177 (Figure 9). The main effect of mission showed that there was not a statistically significant difference in mean C_O between mission, $F(2.917, 78.759) = 2.200$, p = .096, partial η^2 = .075 (Figure 10).

Figure 9
Outdegree Centralization Across Conditions



Note. Error bars indicate 95% CI.

Figure 10
Outdegree Centralization Across Missions



Note. Error bars indicate 95% CI.

These results indicate that teams had similar levels of outdegree centralization between conditions and across missions. This finding suggests that differences in centralization do not appear to manifest for sending messages from a single individual, but rather the centrality of receivers or simply overall activity (senders and receivers) being a larger differentiator between team compositions and task demand levels. C_O does not differentiate between who the sender is, so different roles (navigator, pilot, or photographer) could still have different levels of activity between conditions and missions.

Intensity

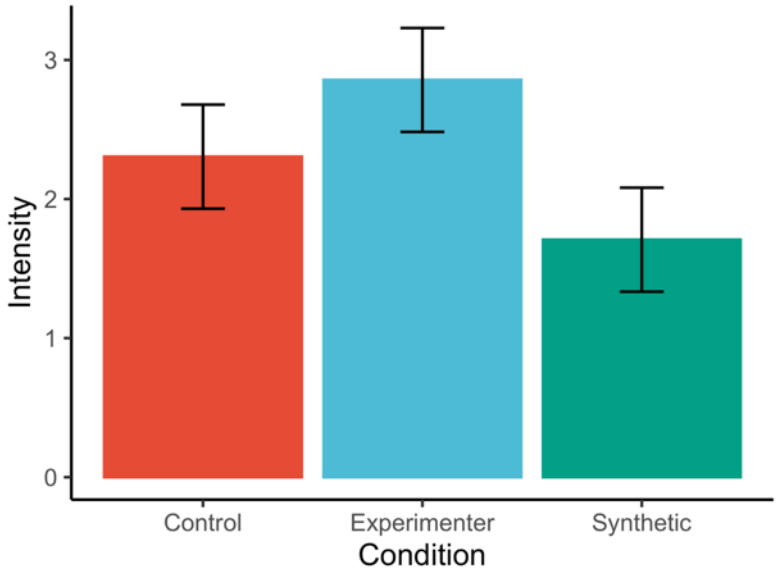
In order to test for differences in communication intensity between conditions and missions, a two-way mixed ANOVA was conducted with intensity as the dependent variable. There were multiple outliers in the data, as assessed by inspection of a boxplot. Normality was violated in Mission 4 for the control condition, and Mission 3 and Mission 4 for the experimenter condition as assessed by Shapiro-Wilk's test ($p < .05$). The decision was made to keep the data intact and not remove outliers or attempt transformation of variables. There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(9) = 21.643$, $p = .010$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

The two-way mixed ANOVA indicated that there was no statistically significant interaction between the condition and mission on intensity, $F(5.897, 79.614) = .572$, $p = .749$, partial $\eta^2 = .041$. The main effect of condition showed a statistically significant difference in mean intensity, $F(2,27) = 9.940$, $p < .001$, partial $\eta^2 = .424$. Pairwise test results (LSD) indicated that the experimenter teams ($M_{Exp} = 2.86$, $SD_{Exp} = .650$) had a

higher mean intensity than control teams (M_{Cont} = 2.31, SD_{Cont} = .942, p = .042) and synthetic teams (M_{Syn} = 1.71, SD_{Syn} = .55, p < .001). The control teams also had higher mean intensity than the synthetic teams (p = .028; Figure 11). The main effect of condition supports the hypothesis that there would be a difference in communication intensity between the different team compositions and follows expectations (Experimenter > Control > Synthetic; H2).

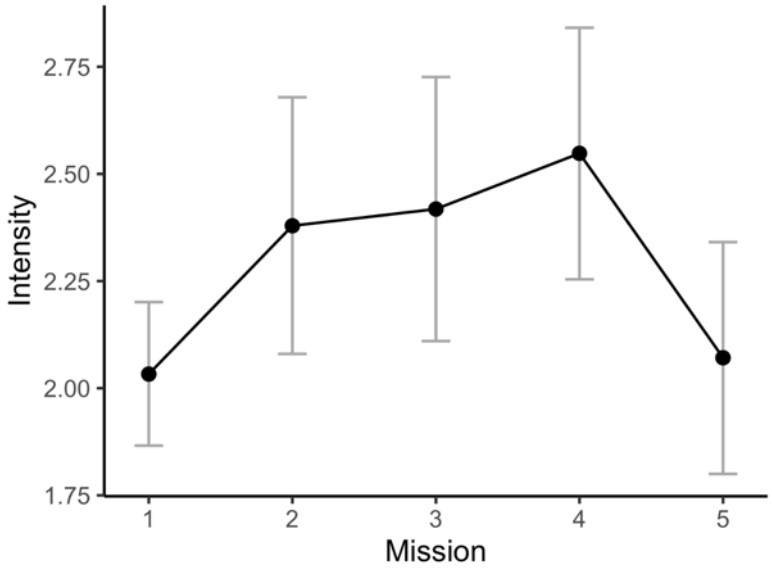
The main effect of mission showed a statistically significant difference in mean intensity, $F(2,27)$ = 9.940, p = .001, partial η^2 = .424 (Figure 12). Pairwise tests results (LSD) indicated that the mean intensity in Mission 2 ($M_{Mission2}$ = 2.379, $SD_{Mission2}$ = .919, p = .013), Mission 3 ($M_{Mission3}$ = 2.418, $SD_{Mission3}$ = .98, p = .007), and Mission 4 ($M_{Mission4}$ = 2.55, $SD_{Mission4}$ = .92, p = .001) was greater than in Mission 1 ($M_{Mission1}$ = 2.03, $SD_{Mission1}$ = .60), and also that intensity in Mission 3 (p = .021) and Mission 4 (p < .001) were greater than Mission 5 ($M_{Mission5}$ = 2.07, $SD_{Mission5}$ = .79). There was no statistically significant difference between intensity in Mission 1 and Mission 5 (p > .05).

Figure 11
Intensity Across Conditions



Note. Error bars indicate 95% CI.

Figure 12
Intensity Across Missions



Note. Error bars indicate 95% CI.

The main effect of the mission partially supports the hypothesis that there would be a reduction in communication intensity in the high task load mission. However, there was not a detectable difference in intensity between Mission 5 and Mission 1 and 2. Reduction in communications during Mission 5 might suggest that higher load from taskwork decreased resource allotment to for communications. An alternate, though not necessarily contradictory explanation for reduced intensity in Mission 5 may be due to a reduction in required communications due to a lower number of targets visited despite that increase in the number of targets available in Mission 5. The lack of an interaction effect between mission and condition suggests that, whereas teams had different overall tendencies for communication intensity, teams appeared to adapt the intensity of their communications across missions in a similar manner under different task load levels.

Performance

Composite Scores and TPE

Team composite and target processing efficiency (TPE) scores have been reported elsewhere (McNeese et al., 2018). In summary for the composite scores, it was found that

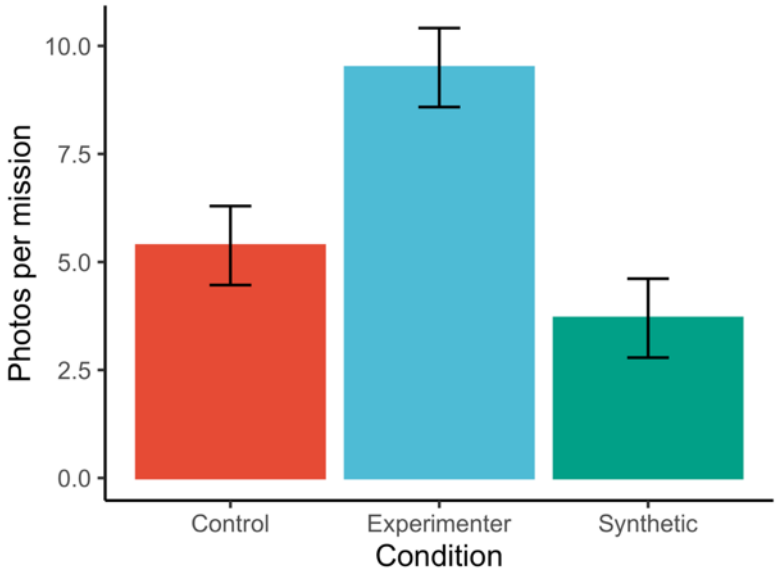
the experimenter teams had higher composite performance scores than the synthetic and control teams, who did not differ significantly from one another. The performance score across missions was not significantly different for Missions 1–4, however, it was found to be significantly lower in Mission 5 for all three conditions. For target processing efficiency, it was found that the experimenter teams had higher TPE than the control teams and synthetic teams, and the control teams were also had higher TPE than the synthetic teams (McNeese et al., 2018).

Target Photos

A two-way mixed ANOVA was conducted to assess for differences in the number of unique good target photos taken per mission. There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(9) = 22.644$, $p = .007$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. The two-way mixed ANOVA indicated that there was not a statistically significant interaction between the condition and mission on the number of photos, $F(5.671, 76.558) = 2.031$, $p = .078$, partial $\eta^2 = .130$. The main effect of condition showed that there was a statistically significant difference in photos between the conditions, $F(2, 27) = 44.866$, $p < .001$, partial $\eta^2 = .769$. Pairwise comparisons (LSD) indicated that the experimenter teams ($M_{Exp} = 9.550$, $SD_{Exp} = 2.10$) took more photos that the control ($M_{Cont} = 5.380$, $SD_{Cont} =$, $p = 2.36$) and synthetic ($M_{Syn} = 3.700$, $SD_{Syn} = 1.90$,$p < .001$) teams, and that the control teams took more photos that the synthetic teams ($p = .013$) (Figure 13). The main effect of mission showed that there was a statistically significant difference in photos taken over missions, $F(2.835, 76.558) = 16.217$, $p < .001$, partial $\eta^2 = .375$. Pairwise comparisons (LSD) indicated that teams took less photos during

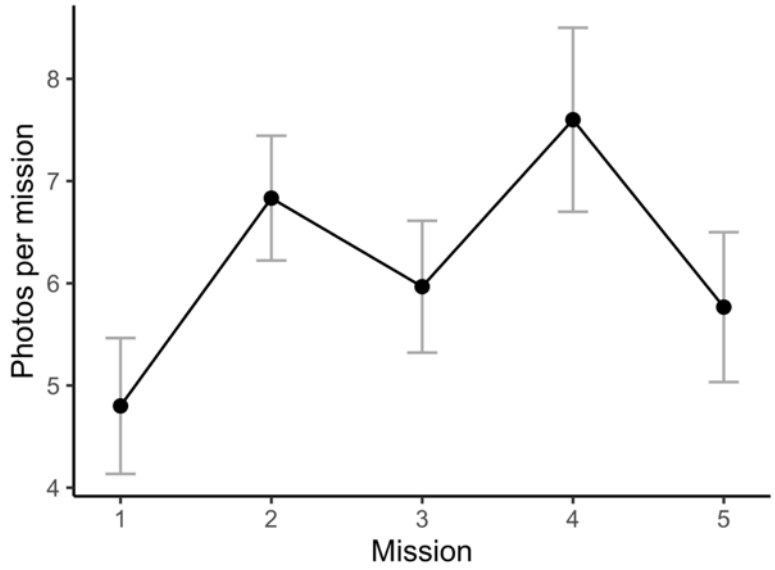
Mission 5 than Mission 2 and Mission 4, but more than Mission 1 and no difference from Mission 3 (Figure 14).

Figure 13
Average Target Photos Across Conditions



Note. Error bars indicate 95% CI.

Figure 14
Number of Target Photos Per Mission Across Missions



Note. Error bars indicate 95% CI.

In summary, the control and synthetic teams had comparable composite scores, but the synthetic teams tended to process targets less efficiently and did not take as many good photos compared to control teams. The experimenter teams performed better overall on all three performance measures.

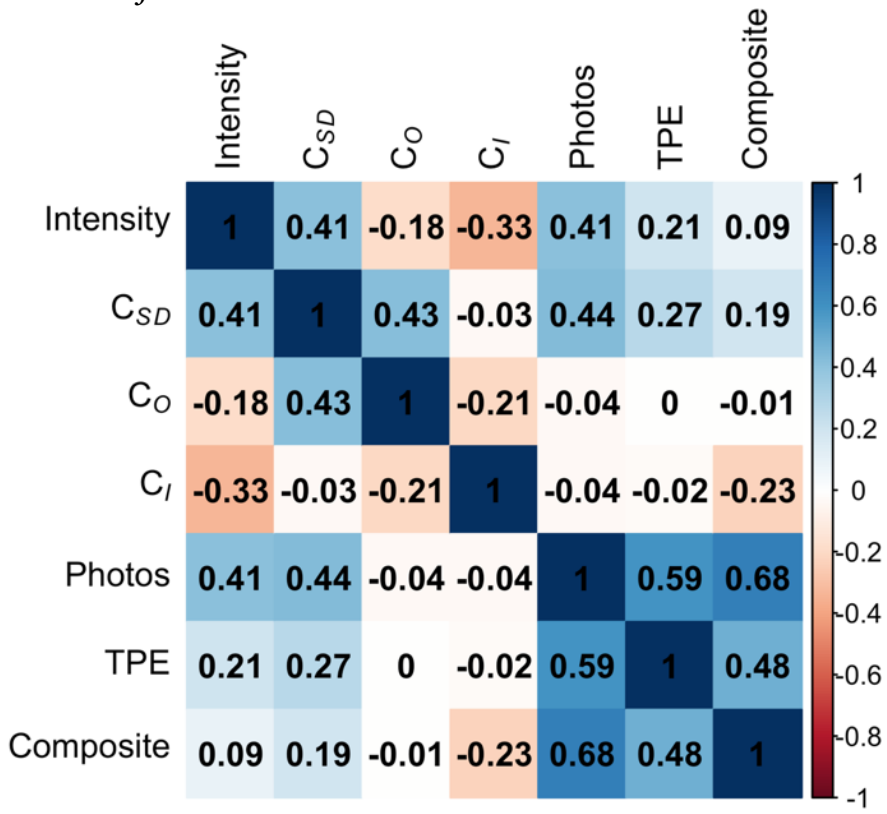
Communications and Performance

Descriptive statistics and linear correlations of communication measures and performance are reported in Table 5 and Figure 15. All three team types (control, experimenter, synthetic) were combined and summarized by mission here. Two of the thirty teams were excluded from this analysis due to missing performance data.

Table 4
Descriptive Statistics of All Observations

	Min	Mean	Max	SD
Intensity	0.68	2.29	5.95	0.87
C_I	0.01	0.18	0.61	0.12
C_O	0.02	0.23	0.56	0.12
C_{SD}	2.00	17.22	47.15	8.77
Photos	0.00	6.19	13.00	3.23
TPE	47.80	863.15	976.54	107.67
Composite	-87.10	314.44	594.16	115.97

Figure 15
Correlation Matrix of All Observations



Note. Includes all observations over conditions and mission with the exception of two teams. C_{SD} = standard deviation of degree centralization, C_O = outdegree centralization, C_I = indegree centralization, TPE = target processing efficiency.

Intensity and Performance. The association between intensity and performance was assessed to test hypothesis H4a, which predicted that increased communication intensity would be related to better team performance. To account for non-independence of repeated measures observations, intensity and performance scores were combined for all five missions and the mean values were calculated for each team. Correlations are displayed in Table 5. There was a significant moderate positive correlation between overall intensity and TPE ($r(28) = .379$, $p = .047$) and also a significant moderate to large positive correlation between intensity and the number of target photos per mission, $r(30) = .500$, $p = .005$. However, intensity was not significantly correlated with composite scores, $r(30) = .138$, $p = .005$.

Table 5
Overall Relationship Between Mean Intensity and Performance

	Intensity	TPE	Composite
TPE	.379*		
Composite	.138	.569**	
Photos	.500**	.753**	.754**

Note. TPE = target processing efficiency, photos = unique good target photos, * = statistically significant at $p < .05$ level, ** = $p < .01$.

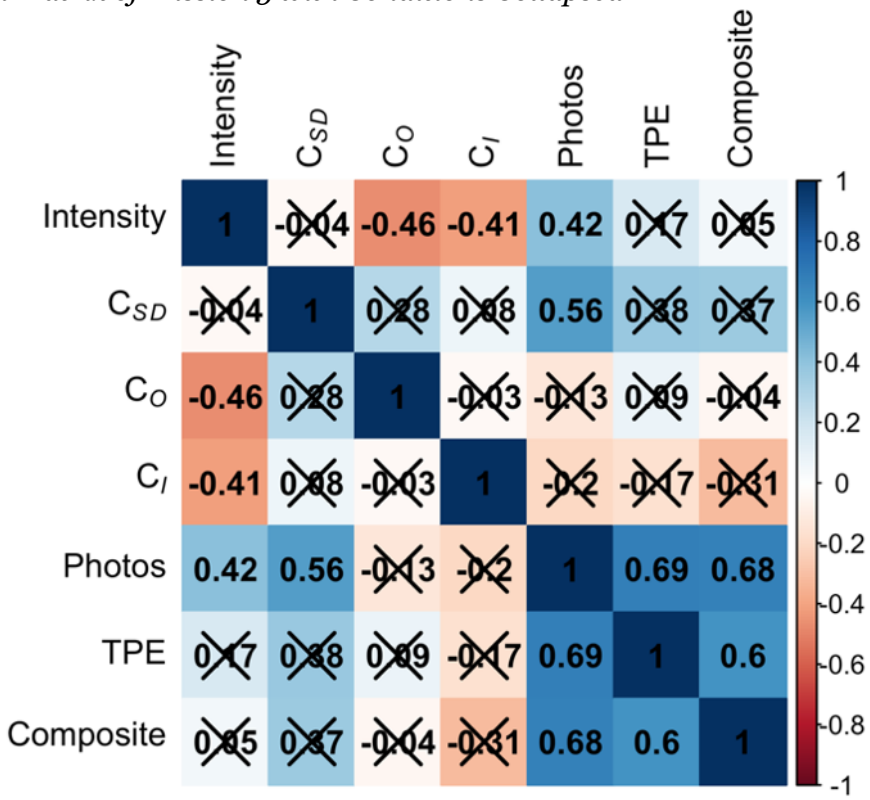
These findings provide partial support for the hypothesis that increased communication intensity would be related with better performance (H4b). However, the direction of causality is not clear, particularly for the increased number of target photos. More photos per mission requires more communication, however, teams who communicated more may have also been able to take more target photos due to better information sharing and coordination.

Improvements in TPE reflect the expectations that better information sharing would improve performance in a task in which coordination is required due to interdependence. However, understanding *when* communications take place could provide more insight, as too much communication when concurrent taskwork was vital (e.g., making flight adjustments or taking a photo) could actually decrease team efficiency during key moments, such as when the aircraft is within a target radius. One explanation for the lack of a correlation between intensity and composite scores is that increased communication intensity may incur communication overhead that results in secondary taskwork penalties which are included in the composite score calculation (Appendix B). This effect may have been enough to offset the benefits of increased intensity in this task, but not so much as to result in a negative correlation between composite scores and intensity.

Centralization and Performance

In order to evaluate the hypothesis that reduced centralization would be related to performance in under higher task demand in Mission 5 (H3b), linear correlations were examined for mission 5 alone and tested for statistical significance (Figure 16).

Figure 16
Correlation Matrix of Mission 5 with Conditions Collapsed



Note. C_{SD} = standard deviation of degree centralization, C_O = outdegree centralization, CI = indegree centralization, TPE = target processing efficiency. Non-significant correlations ($p > .05$) are crossed out.

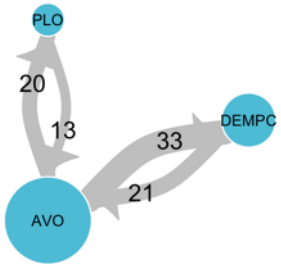
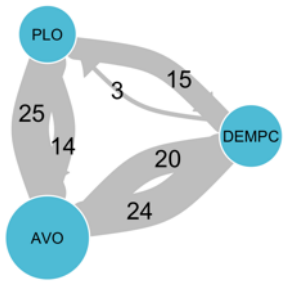
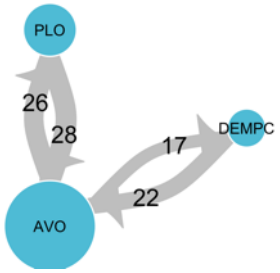
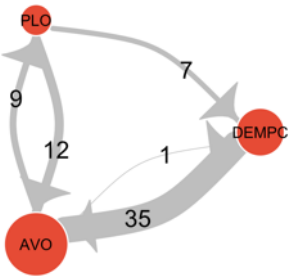
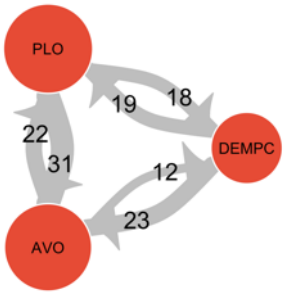
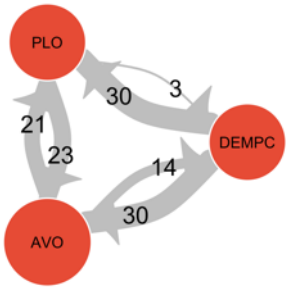
Only C_{SD} was significantly positively correlated with photos per mission ($p < .05$), all other measures of centralization failed to reach significance. This did not support the hypothesis that decreased centralization would be associated with performance in mission 5 (H2b). Once again, intensity had an interesting moderate negative correlation with C_O and C_I.

High and Low Performers

Three top performing teams and three bottom performing teams were selected according to their composite scores in Mission 5 for exploratory analysis and are summarized in Table 6. Somewhat surprisingly, two of the three high performing teams (30 and 22) did not have any communication between the navigator and photographer at all during Mission 5. This implies that when a successful target photo was taken, feedback indicating movement to the next waypoint was provided from the photographer to the pilot, but not to the navigator (See Figure 2). It is plausible that the navigators relied on observations implicit coordination cues such as the pilot queuing of the next waypoint as a form of feedback, rather than explicit communication from the navigator. In contrast, all three low performing teams, who exhibited varying amounts of communication between the photographer and navigator. The high performers tended to have a moderate level of intensity, all three of which were within one standard deviation of the mean of all observations (M = 2.99). The lower performers had values further from the overall average—either much higher or much lower. Team 9 exhibited an extreme example of both low intensity and highly centralized communications, both of which could be expected to be associated with poor performance. All three high performing teams had an experimenter in the pilot roles, and all three low performers had a naïve participant in the pilot role.

Table 6

High and Low Performing Teams in Mission 5

	Team 30	**Team 29**	**Team 22**
High performers			
Cond.	Experimenter	Experimenter	Experimenter
Intensity	2.18	2.53	2.33
C_{SD}	27.22	13.80	27.87
C_I	0.09	0.06	0.31
C_O	0.41	0.17	0.19
Comp.	473.21	430.44	420.18
TPE	974.36	969.64	970.35
Photos	12	10	11
	Team 9	**Team 25**	**Team 8**
Low performers			
Cond.	Control	Control	Control
Intensity	1.60	3.13	3.03
C_{SD}	14.50	9.87	6.35
C_I	0.60	0.15	0.16
C_O	0.32	0.09	0.24
Comp.	74.48	157.47	165.98
TPE	875.45	855.86	800.16
Photos	5	6	2

Workload Distribution

The relationship between workload distribution and centralization in the high demand mission was explored by examining differences in the subjective workload of teammates. Subjective workload ratings were collected from participants after Mission 1 (baseline) and Mission 5 (high demand) using a modified 15-question NASA-TLX (Appendix C; Hart & Staveland, 1988). Questions were split into 6 dimensions: mental

demand, physical demand, temporal demand, effort, performance, and frustration and responses were provided on a 5-point Likert scale. Because the role of the pilot was filled by an experimenter and a synthetic agent in two of the conditions, only the responses from the photographer and navigator were considered for this analysis. Mean responses to all 15 questions were computed for each role. The absolute value of the difference in NASA-TLX responses between the two teammates was computed for each team to produce a single score approximating the workload distribution. With this approach, a team with a high navigator score and a low photographer score would be considered the same as if their responses were swapped. A two-way mixed ANOVA was conducted to assess for difference between sessions and conditions. No significant interaction or main effects for condition and mission were found ($p > .05$).

Next, correlations were analyzed between team score distributions and measures of centralization during the high load mission (Mission 5). There was a statistically significant, moderate positive correlation between C_O in Mission 5 and the absolute difference in NASA-TLX responses recorded immediately following the mission, $r(30) = .445$, $p < .014$ (Table 7). No other statistically significant correlations were found for session 1 or session 2. This lends partial support to the hypothesis that increased centralization would be related to larger differences in individual workload distributions (H5).

Table 7

Correlation of Centralization and team NASA-TLX Differences Following Mission 5

	TLX Difference	C_O	C_I
C_O	.445*		
C_I	-.104	-.052	
C_{SD}	-.049	2.73	.067

Note. C_{SD} = standard deviation of degree centrality, C_I = indegree centralization, C_O = outdegree centralization, * = statistically significant at $p < .05$ level.

Summary of Results

Overall findings are summarized by hypothesis in Table 8.

Table 8
Summary of Main Findings by Hypothesis

	Hypothesis	Findings
1	Difference in centralization between conditions: Exp > Cont = Syn	C_{SD}: Supported; Exp > Syn = Cont C_I: Not supported; Cont > Syn = Exp C_O: Not supported; Cont = Syn = Exp
2	Difference in intensity between conditions: Exp > Cont > Syn	Supported; Exp > Cont > Syn
3a	Lower centralization in high demand Mission 5 than in Missions 1–4	C_{SD}: Partially supported; Mission 5 < 2, 3, 4 C_I: Not supported; Mission 5 = 3, 4 C_O: Not supported Mission 5 = 1, 2, 3, 4
3b	Negative correlation between centralization and performance in Mission 5	C_{SD}: Not supported C_I: Not supported C_O: Not supported
4a	Lower intensity in Mission than in Missions 1–4	Partially supported; Mission 5 < 3, 4
4b	Positive correlation between intensity and performance	Partially supported; positive correlations of intensity with TPE and photos per target, but not composite
5	Centralization positively correlated a larger difference in individual workload distribution in mission 5	Partially supported; positive correlation between C_O and differences in NASA-TLX ratings with teams

CHAPTER 6

DISCUSSION

Summary

This thesis was motivated by two primary research questions. First, it set out to explore how a team's communication network is associated with team workload and performance. Second, it sought to examine how communication networks between human–autonomy teams and two types of all-human teams differ under high task demand. These research questions were explored by examining the relationship between team workload and communications in a small command and control team task using a social network analysis approach. Three different team compositions (all-human participant teams, all-human teams with a highly trained experimenter, and human–autonomy teams) worked together in the Remotely Piloted Aerial System synthetic task environment (RPAS-STE) over a series of five missions under different levels of task demand. Text communications were recorded, and measures of communication centralization and intensity were computed to assess differences between the different team compositions and demand levels, and team workload distributions were assessed using NASA-TLX.

Results of this study demonstrated some differences in communications between conditions and levels of task demand, including some partial support for a reduction in centralization and intensity in the high demand mission. However, only one of the three performance measures (number of photos) was associated with decentralized communications (C_{SD}) under high demand. Increased communication intensity was positively correlated with two of the three performance measures across all missions (number of photos and target processing efficiency). Partial support was provided to the hypothesis that higher centralization would result in uneven workload distributions.

Overall this study provides insight into how team human–autonomy and all-human team communication networks change under load in a command and control task environment and explored the use of SNA methods in this context. However further exploration of the relationship between centralization, intensity, and performance is needed.

Discussion of Findings

It was hypothesized that the centralization of communications would be different between control, experimenter, and synthetic teams. Experimenter teams were expected to have more centralized communications than teams in the control and synthetic condition. This hypothesis was supported when the analysis was restricted to C_{SD}, a measure of non-directional network heterogeneity adapted for the small weighted networks used in this study and used to indicate centralization. However, C_I, another measure of centralization which only considers incoming messages, indicated that the control condition had higher centralization than the synthetic and experimenter conditions. These results mean that experimenter teams tended to have higher overall differences in activity between the activity of team members, whereas control teams tended to centralize sending messages to a single individual. Due to the nature of the text communication system in this task, it is possible that receiving a large quantity of messages had less of an impact on individual workload than sending them, thus the tendency for control teams to have higher indegree centralization, but lower intensity and C_{SD}. Text messages were stored by the receiver's chat interface, were retrieved on demand, and could be referenced again later if necessary. Sending a message, on the other hand, occupied the sender's visual attention at the time of sending. However, the trained experimenter in the experimenter condition was likely to have developed more skill at interleaving sending and receiving of messages with other tasks, minimizing

interruptions, and enabling higher overall pilot message activity (incoming and outgoing), which could result in a higher C_{SD} for the team. Further analysis of individual node-level measures is required to test these hypotheses. One implication for this finding is that different measures of centralization are likely to have different implications for team workload when communication systems resource requirements (e.g., audio vs. visual, sending vs. receiving) is considered. Future work could define links between nodes using multiple dimensions characterizing team interactions (e.g., Johnson et al., 2020) to determine impacts of different communication media within a network.

The more decentralized nature of the synthetic teams' communications suggest that either (1) the human participants (navigator and photographer) tended to communicate with one another proportionally more in that condition compared to the other two, (2) that the pilot had a lower communication frequency, or (3) a combination or both. The limited language production and comprehension capabilities of the synthetic agent make the first explanation plausible. However, this also suggests team workload may manifest in different ways in human–autonomy teams. On one hand, an artificial agent may be able to contribute capabilities that are not possessed by any one member of the team, such as executing large computations quickly and being immune to the effects of fatigue. However, teammates may not rely on the artificial agent to act appropriately when conditions are less than ideal such as peak workload periods and complex situations. If decentralization of communications—and presumably redistribution of taskwork—is a real phenomenon in teams under high load as has been suggested by previous work, then swapping out a human teammate for an artificial agent may in some cases result in reducing a team's workload capacity. For example, if tasks must be quickly reallocated in a team to meet increasing demands, the rigid nature of artificial agents may limit their ability to provide assistance if the tasks are outside of the

agent's performance envelope. Although this is also the case for human operators who do not have access to the requisite skills or tools, the ability of humans to develop novel solutions may still place them at an advantage over an artificial agent in such a situation. Whereas these types of limitations on the capacity of a team to reallocate task is not normally considered a workload limitation, bottlenecks, delays, and sub-optimal workload distribution can result from such a limitation and actually be the manifestation of reduced team workload capacity. Therefore, it may be prudent to first consider s*upplementing* teams with artificial agents, rather than attempting to outright replace humans in existing command and control systems that need to function under high load and complexity. Iterative testing and evaluation of automated systems embedded into teams should take place in both routine and especially under peak workload conditions, and consideration should be made of how team communication networks transform under high demand.

Overall, higher intensity team communications were associated with better target processing efficiency and more photos taken per mission, aligning with similar work in larger organizations and teams suggesting that more intense and dense communication networks can improve team performance (Balkundi & Harrison, 2006; Grund, 2012; Mora-Cantallops & Sicilia, 2019; Sparrowe et al., 2001). However, this positive relationship was not detected with composite scores. Composite score calculations penalized poor performance on secondary tasks such as aircraft system monitoring and resource management, whereas TPE captured the efficiency and performance of the primary team task: coordinating as a team to acquire good photos of targets. It is possible that teams may have made sacrifices (consciously or not) in their secondary taskwork to prioritize communicating and executing the primary team task, similar to the task shedding under high workload suggested in other studies (Wickens & Tsang,

2015). Further exploration of this discrepancy could include examination of individual secondary task performance to assess whether tradeoffs were made between primary taskwork, secondary taskwork, and teamwork activities under different levels of task demand.

The higher intensity of communications in the experimenter condition combined with the findings that they performed better overall, suggest that those teams may have had a greater capacity to balance teamwork and taskwork demands, had more effective communications overall, and therefore less team-internal data limitations. Two primary factors likely contributed to this. The trained experimenter almost certainly had a higher skill level in conducting piloting taskwork than the majority of naïve participants, resulting in relatively less individual workload for the pilot, more efficient movement of the RPA between targets, and a reduction in bottlenecking. Secondly, the trained experimenter pushed and pulled essential information in a timely manner, which previous findings suggest positively influenced coordination behaviors of the entire team (Demir et al., 2017b; Johnson et al., *in press*). Both of these may have increased the team's workload capacity by improving both the efficiency of information flow through the team, and the timing of activities —reducing redundancies, working at cross purposes, and desynchronized taskwork.

Timely pushing of information has also been associated with anticipating the informational needs of others, and more generally good teamwork in this task (Gorman et al., 2006). Previous analysis of this dataset found that synthetic teams tended to push information to one another less often in general, which led to more pulling behaviors when compared to the higher performing experimenter teams (Demir et al., 2017). Pushing may indicate the development of more effective implicit coordination strategies over time, which can help teams operate effectively in some high workload situations by

reducing coordination overhead (Entin & Serfaty, 1999; MacMillan et al., 2004). On the simplest level, a push only requires one message, whereas a pull requires two at a minimum (request and response). The reduction in communication intensity in Mission 5 was found across teams and lends support to the hypothesis that the high demand mission led to a reduction in communications. Future work could consider how pushing and pulling of information (e.g., "anticipation ratio"; Entin & Serfaty, 1999) interacts with network characteristics such as centralization and intensity.

Teams were hypothesized to decentralize during the high demand mission compared to the other missions in order to adapt to the increased load. This study found that teams did decentralize during the higher demand mission when compare to the lower demand missions (except for the first) when using C_{SD} as a measure of decentralization, but C_I showed an increase in centralization in the high demand mission, and no significant difference in C_O. An overall decrease in centralization would aligns with previous research demonstrating that teams tend to decentralize their communications under increased task demands and complexity (Shaw et al., 1965; Brown & Miller, 2000; Katz, Lazer, Arrow, & Contractor, 2004; Barth, Schraagen, & Schmettow, 2015). However, the results of this study are mixed. Somewhat surprisingly, there were not a detectable difference in decentralization between the different conditions when comparing communications in the high demand mission. This implies that all three team types decentralized their communications in a similar manner to the extent that they did change their communication structure. This is interesting considering the substantial difference between the three team compositions.

Teams that decentralized during the high demand mission were also expected to perform better. However, decentralization was not found to be associated with better performance in the high demand mission. This may be explained by a closer examination

of the task. In this experiment, higher task demand was characterized by increased information processing requirements, but not necessarily a significant increase in complexity, uncertainty, or a clear need for *novel* integration of heterogeneous information. That is, teams needed to plan for more targets, with more route restrictions, and encountered more roadblocks. However, the nature of their taskwork did not change significantly. Increased complexity is not necessary for increasing team workload. However, a higher information processing requirement (or reduction in the team's workload capacity) should increase team workload. Increasing demand in this task could be accomplished by either increasing the rate at which the team moves between targets or increasing the number of targets that are encountered at a given airspeed. However, teams in this task could self-regulate the pace of their performance by adjusting airspeed within restrictions, rather than having the task demand (processing rate) imposed on them externally. This meant that they could slow the task demand to a certain extent, and scale it to their capacity to execute tasks. The primary differences in task demand for Mission 5 were increased roadblocks, increased target availability, and slightly less distance between waypoints when compared to other missions. In order to overcome a roadblock in this experiment, replanning was required to take pictures of *ad hoc* targets. This activity may have been too similar to routine performance, rather than truly novel circumstances that required unique coordination and task reallocation and redistribution. Furthermore, backup behaviors were generally not possible due to the constraints of the task environment (e.g., the navigator couldn't do the photographer's job if they were overloaded), and the capacity to redistribute taskwork was limited. The implication is that if tasks are complex and require integration of heterogeneous information, or if tasks put specific individuals at risk of overload, systems should be designed to enable both rapid and distributed information sharing and the capacity to

assist in taskwork when necessary. This may enable teams to expand their workload capacity through decentralization when task load reaches peak levels.

This thesis also examined the relationship between centralization and workload distribution. Highly centralized teams may be bottlenecked, leading to delays in information flow through an individual, which could manifest as uneven workload distribution. This hypothesis was only partially supported by the correlation between differences in teammate NASA-TLX scores and C_O in Mission 5. However, other measures of centralization were not correlated with workload distributions. Further examination is required to identify bottleneck effects and improve confidence in the potential finding. Future assessments of bottleneck effects could benefit from increased frequency of workload assessment and the use of other measures, such as a secondary task paradigm. Also, certain contexts and task types may be more prone to bottlenecks than others. For example, in Roberts et al. (2017; See also Stanton et al., 2020), a communication bottleneck was identified a submarine control room where a single individual (the sonar controller) was the conduit through which six individuals on the team shared information. This was due to the physical layout of the control room, the assignment of tasks, and the design of communication networks. Modifications to those constraints, including co-locating key specific roles, alleviated the bottleneck and were associated with more efficient communication and task completion. The RPAS-STE on the other hand does not require information transfer through an intermediate teammate during routine conditions, and therefore bottleneck effects are likely to be limited in this task.

Applications.

Interaction-based measures of teamwork continue to have potential for allowing human operators and artificial agents to collaborate more effectively despite their vastly

differing cognitive capacities. In future systems, communication-network derived team workload metrics could provide feedback to other operators or artificial agents in real time, improving awareness for interruptions, backup behaviors, and other coordination considerations. Team communication network properties can also be used as a means to anticipate workload bottlenecks in future system design. In the training context, evaluation of network properties could be used be used to examine and even compare network structures between teams, such as between high and low performing teams.

Limitations

There were a number of limitations for this study. Although social network analysis provides a number of advantages when compared to other approaches to measuring interaction in teams, it also has limitations. First, the number of team members impacts the applicability and interpretation of measures. For example, the small social networks utilized in this thesis limited the availability of centrality measures to variations of degree centrality, and different centrality measures can have vastly different interpretations (Borgatti, 2005). Larger teams (e.g., 5 or more) would lend themselves more readily to other measures of centrality. For example, variations of *betweenness* and *closeness* centrality may be another way of assessing bottlenecks (Freeman, 1978; Stanton & Roberts, 2017). Furthermore, caution should be practiced when generalizing SNA measures between different sized networks, as the interpretation of some measures may not scale to larger teams and multi-team systems (Borgatti et al., 2018). Static SNA methods have the advantage of providing a systems-level view of aspects of teamwork, however, they do not capture the dynamics of how teamwork changes over time. Other approaches such as dynamic network analysis may provide further insight that can supplement the SNA toolkit (Carley, 2003).

Other limitations are related to the study design for answering questions about team workload. Some differences were found in communication behaviors between team compositions and missions of varying task demand, however the link between resource allocation and performance for this task is not completely clear. The highly heterogenous nature of the roles meant that backup behaviors were limited. For example, the photographer could not be assisted in physically adjusting the camera settings and taking a photo, and no one except for the pilot could manipulate the flight controls. This potentially exaggerated role heterogony in the RPAS-STE is extremely useful for answering questions related to communication, coordination, situation awareness, and team cognition (Cooke, 2015). However, more flexible and overlapping roles—such as the capacity to take over aspects of a teammate's job if needed—would allow for more observation of both top-down and bottom-up task reallocation as well as backup behaviors, both of which are important aspects of team workload.

Another area of difficulty was the quantification of workload and task demands. Mission 5 was considered high demand relative to missions 1–4 and this manipulation of task difficulty is important for interpreting aspects of the within-subjects and interaction effects in the study design. However, as previously mentioned, all 4 missions were partially self-paced by the team, which allowed teams to self-regulate and possibly avoid real overload situations. For example, by slowing down their airspeed to the lower end of the airspeed restrictions. This seems especially likely in the control condition where there is not a scripted confederate or artificially agent influencing the pace of the participants' coordination. Future experiments examining team workload with a task demand variable should consider how the study design will induce higher workload and overload.

Conclusion

Traditional approaches to command and control include unitary decision rights, hierarchical patterns of interaction, and limited distribution of information. Advances in information technology have enabled organizations to adopt more decentralized structures that are expected to be better situated to evolve, adapt, and self-organize in a way which allows their members to perform complex tasks in distributed space (Stanton et al., 2012). These decentralized "teams of teams" (McChrystal et al., 2015) can be expected to become more common in future environments in which resources and organizational components may be temporally and geographically dispersed, but coordination is still required to achieve collective focus, convergence and agility (Alberts, 2007; Woods & Branlat, 2010). These teams can also be expected to include more autonomous technology teaming with human counterparts (Knotts & Alberts, 2017). It is important to understand how workload emerges at the team-level and as a result of the interaction of individuals, systems, and tool in these environments. This thesis applied concepts from social network analysis in a novel manner to better understand team workload and team communications in a small command and control team task, with a special focus on understanding how artificially intelligent agents might have an impact on teamwork. Differences were uncovered between communications in all-human and human–autonomy teams, with some preliminary evidence for workload bottlenecks associated with centralization, but evidence for link between communication decentralization and performance under load in this task was not clear cut. This thesis may help generate hypotheses and design considerations for future experiments in a similar context. Future work should continue to pursue an understanding of team workload and communications in larger teams and complex sociotechnical systems.

REFERENCES

Alberts, D. S. (2007). *Agility, focus, and convergence: The future of command and control*. Office of the Assistant Secretary of Defense for Networks and Information Integration Washington DC.

Argote, L., Turner, M. E., & Fichman, M. (1989). To centralize or not to centralize: The effects of uncertainty and threat on group structure and performance. *Organizational Behavior and Human Decision Processes*, *43*(1), 58–74. https://doi.org/10.1016/0749-5978(89)90058-7

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129–135). Elsevier.

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal*, *49*(1), 49–68.

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.

Barabási, A.-L., & Frangos, J. (2002). *Linked: The new science of networks science of networks*. Basic Books.

Barocas, A., Ilany, A., Koren, L., Kam, M., & Geffen, E. (2011). Variance in Centrality within Rock Hyrax Social Networks Predicts Adult Longevity. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0022375

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, *101*(11), 3747–3752.

Barth, S., Schraagen, J. M., & Schmettow, M. (2015). Network measures for characterizing team adaptation processes. *Ergonomics*, *58*(8), 1287–1302.

Bedwell, W. L., Salas, E., Funke, G. J., & Knott, B. A. (2014). Team workload: A multilevel perspective. *Organizational Psychology Review*, *4*(2), 99–123.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks*. Sage.

Bowers, C. A., Braun, C. C., & Morgan Jr, B. B. (1997). Team workload: Its meaning and measurement. In *Team performance assessment and measurement* (pp. 97–120). Psychology Press.

Bowers, C. A., & Jentsch, F. (2005). Team Workload. In N. A. Stanton, K. Brookhuis, E. Salas, & H. Hendrick, *Handbook of human factors and ergonomics methods*. CRC Press.

Bowers, C. A., Urban, J. M., & Morgan Jr, B. B. (1992). *The study of crew coordination and performance in hierarchical team decision making* (No. TR-92-01). University of Central Florida.

Broadbent, D. E. (1958). *Perception and communication*. Elsevier.

Broadhurst, P. L. (1957). Emotionality and the Yerkes-Dodson law. *Journal of Experimental Psychology*, *54*(5), 345.

Brown, T. M., & Miller, C. E. (2000). Communication networks in task-performing groups: Effects of task complexity, time pressure, and interpersonal dominance. *Small Group Research*, *31*(2), 131–157.

Carley, K. M. (2003). *Dynamic network analysis*. Center for Computational Analysis of Social and Organizational Systems, Carnegie Mellon University, Pittsburgh, USA.

Christopher Wickens. (2000). Wickens, C. D. (2000). The trade-off of design for routine and unexpected performance: Implications of situation awareness. In *Situation Awareness Analysis and Measurement* (pp. 211–225). CRC Press.

Cooke, N. J. (2005). Moderating effects of team role on workload demands of teams. *Human Performance, Situation Awareness, and Automation: Current Research and Trends HPSAA II, Volumes I and II*, 235.

Cooke, N. J. (2015). Team cognition as interaction. *Current Directions in Psychological Science*, *24*(6), 415–419.

Cooke, N. J., Demir, M., & McNeese, N. (2016). *Synthetic Teammates as Team Players: Coordination of Human and Synthetic Teammates* (RE2016844 01). Cognitive Engineering Research Institute Mesa United States. https://apps.dtic.mil/docs/citations/AD1017169

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

Cooke, N. J., Gorman, J. C., Winner, J. L., & Durso, F. T. (2007). Team cognition. *Handbook of Applied Cognition*, *2*, 239–268.

Cooke, N. J., Gorman, J., Pedersen, H., Winner, J., Duran, J., Taylor, A., Amazeen, P. G., Andrews, D. H., & Rowe, L. (2007). *Acquisition and Retention of Team Coordination in Command-and-Control*. Cognitive Engineering Research Institute, Mesa, AZ. https://apps.dtic.mil/docs/citations/ADA47556

Cooke, N. J., Shope, S. M., Schiflett, L. R. E., Salas, E., & Coovert, M. D. (2004). Designing a synthetic task environment. *Scaled Worlds: Development, Validation, and Application*, 263–278.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695.

Cummings, M. L., Clare, A., & Hart, C. (2010). The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling. *Human Factors*, *52*(1), 17–27. https://doi.org/10.1177/0018720810368674

Cummings, M. L., & Guerlain, S. (2007). Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles. *Human Factors*, *49*(1), 1–15. https://doi.org/10.1518/001872007779598109

Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2018). Team coordination and effectiveness in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems*, *49*(2), 150–159.

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3–12.

Demir, M., McNeese, N. J., & Cooke, N. J. (2019). The Evolution of Human-Autonomy Teams in Remotely Piloted Aircraft Systems Operations. *Frontiers in Communication*, *4*. https://doi.org/10.3389/fcomm.2019.00050

Dias, R. D., Zenati, M. A., Stevens, R., Gabany, J. M., & Yule, S. J. (2019). Physiological synchronization and entropy as measures of team cognitive load. *Journal of Biomedical Informatics*, *96*, 103250.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*(1), 5–27.

Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41(2), 312–325. https://doi.org/10.1518/001872099779591196*

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Funke, G. J., Knott, B. A., Salas, E., Pavlas, D., & Strang, A. J. (2012). Conceptualization and measurement of team workload: A critical need. *Human Factors*, *54*(1), 36–51.

Gopher, D., & Donchin, E. (1986). *Workload: An examination of the concept*.

Gopher, D., & Donchin, E. (1986). *Workload: An examination of the concept*. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2. Cognitive processes and performance* (p. 1–49). John Wiley & Sons

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences*, *14*(3), 265.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, *49*(12–13), 1312–1325.

Gorman, J. C., Demir, M., Cooke, N. J., & Grimm, D. A. (2019). Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: a layered dynamics approach. *Ergonomics*, *62*(5), 629-643.

Grayson, T. (2018). *Mosaic Warfare*. https://www.darpa.mil/attachments/STO-Mosaic-Distro-A.pdf

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008). The red-line of workload: Theory, research, and design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(18), 1204–1208.

Grimm, D., Demir, M., Gorman, J. C., & Cooke, N. J. (2018). Systems Level Evaluation of Resilience in Human-Autonomy Teaming under Degraded Conditions. *2018 Resilience Week (RWS)*, 124–130. https://doi.org/10.1109/RWEEK.2018.8473561

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, *8*(3), 483–500. https://doi.org/10.1075/is.8.3.10gro

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, *34*(4), 682–690.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining Time and Intensity Effects in Assessing Operator Information-Processing Load. *Human Factors*, *39*(1), 30–47. https://doi.org/10.1518/001872097778940597

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*(2), 174–196. https://doi.org/10.1145/353485.353487

Houghton, R. J., Baber, C., Cowton, M., Walker, G. H., & Stanton, N. A. (2008). WESTT (workload, error, situational awareness, time and teamwork): An analytical prototyping system for command and control. *Cognition, Technology & Work*, *10*(3), 199–207. https://doi.org/10.1007/s10111-007-0098-4

Hughes, A. M., Hancock, G. M., Marlow, S. L., Stowers, K., & Salas, E. (2019). Cardiac Measures of Cognitive Workload: A Meta-Analysis. *Human Factors*, *61*(3), 393–414. https://doi.org/10.1177/0018720819830553

Hutchins, E. (1995a). *Cognition in the Wild*. MIT press.

Hutchins, E. (1995b). How a Cockpit Remembers Its Speeds. *Cognitive Science*, *19*(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1

Johnson, C. J., Demir, M., Wolff, A., & Cooke, N. J. (In press). The impact of team training on coordination and trust calibration in human–autonomy teaming.

Johnson, C. J., Lematta, G. J., Huang, L., Holder, E., Bhatti, S., & Cooke, N. J. (2020, July). An Interaction Taxonomy of Human–Agent Teaming in Next Generation Combat Vehicle Systems. In International Conference on Applied Human Factors and Ergonomics (pp. 10-16). Springer, Cham.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction*, *3*(1), 43–69.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063).

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*(4), 657–690. https://doi.org/10.1037/0021-9010.74.4.657

Katz, N., Lazer, D., Arrow, H., & Contractor, N. (2004). Network Theory and Small Groups. *Small Group Research*, *35*(3), 307–332. https://doi.org/10.1177/1046496404264941

Klien, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(6), 91–95. https://doi.org/10.1109/MIS.2004.74

Kott, A., & Alberts, D. S. (2017). How Do You Command an Army of Intelligent Things? *Computer*, *50*(12), 96–100. https://doi.org/10.1109/MC.2017.4451205

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology*, *46*(1), 38–50. https://doi.org/10.1037/h0057189

Lee, P.-J., Kolling, A., & Lewis, M. (2011). Workload Modeling using Time Windows and Utilization in an Air Traffic Control Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 846–850. https://doi.org/10.1177/1071181311551176

Lematta, G. J., Johnson, C. J., Chiou, E. K., & Cooke, N. J. (2019). Does Team Interaction Exploration Support Resilience in Human Autonomy Teaming? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 1866–1866. https://doi.org/10.1177/1071181319631492

MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: The hidden cost of team cognition. *Team Cognition: Process and Performance at the Interand Intra-Individual Level. American Psychological Association, Washington, DC. Available at Http://Www. Aptima. Com/Publications/2004_MacMillan_EntinEE_Serfaty. Pdf.*

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, *26*(1), 87–119. https://doi.org/10.1145/174666.174668

McChrystal, G. S., Collins, T., Silverman, D., & Fussell, C. (2015). *Team of teams: New rules of engagement for a complex world*. Penguin.

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.

Mitchell, D. (2000). *Mental Workload and ARL Workload Modeling Tools*. Army Research Lab Aberdeen Proving Ground MD.

Mora-Cantallops, M., & Sicilia, M.-Á. (2019). Team efficiency and network structure: The case of professional League of Legends. *Social Networks*, *58*, 105–115.

Myers, C., Ball, J., Cooke, N., Freiman, M., Caisse, M., Rodgers, S., Demir, M., & McNeese, N. (2018). Autonomous Intelligent Agents for Team Training. *IEEE Intelligent Systems*, *34*(2), 3–14.

Neville A. (Neville Anthony) author. Stanton. (2018). *Systems Thinking in Practice Applications of the Event Analysis of Systemic Teamwork Method*. CRC Press LLC.

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, *70*(5), 056131. https://doi.org/10.1103/PhysRevE.70.056131

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64. https://doi.org/10.1016/0010-0285(75)90004-3

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Parks, D., & Boucek, G. (1988). Workload Prediction, Diagnosis, and Continuing Challenges. In G. McMillan, D. Beevis, E. Salas, M. Strub, R. Sutton, & L. Van Breda, *Applications of Human Performance Models to System Design* (pp. 47–63). Plenum Press.

Patterson, E. S., Watts-Perotti, J., & Woods, D. D. (1999). Voice loops as coordination aids in space shuttle mission control. *Computer Supported Cooperative Work (CSCW)*, *8*(4), 353–371.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Roberts, A. P. J., Stanton, N. A., Fay, D. T., & Pope, K. A. (2019). The effects of team co-location and reduced crewing on team communication characteristics. *Applied Ergonomics*, *81*, 102875. https://doi.org/10.1016/j.apergo.2019.102875

Roscoe, A. H. (1984). *Assessing Pilot Workload in Flight*. Royal Aircraft Establishment, Bedford, United Kingdom. https://apps.dtic.mil/docs/citations/ADP004109

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In *Teams: Their training and performance* (pp. 3–29). Ablex Publishing.

Salas, E., & Fiore, S. M. (Eds.). (2004). *Team cognition: Understanding the factors that drive process and performance* (pp. xi, 268). American Psychological Association. https://doi.org/10.1037/10690-000

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). *Automation Surprises*.

Schmidt, D. K. (1978). Queueing analysis of the air traffic controllers workload. *IEEE Transactions on Systems, Man and Cybernetics*, Article VSMC-N6. https://trid.trb.org/view/80198

Sellers, J., Helton, W. S., Näswall, K., Funke, G. J., & Knott, B. A. (2014). Development of the team workload questionnaire (TWLQ). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 989–993.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of information. *Urbana: University of Illinois Press*, *97*.

Shaw, M. E. (1964). Communication Networks. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 1, pp. 111–147). Academic Press. https://doi.org/10.1016/S0065-2601(08)60050-7

Snijders, T. A. B. (1981). The degree variance: An index of graph heterogeneity. *Social Networks*, *3*(3), 163–174. https://doi.org/10.1016/0378-8733(81)90014-9

Sparrowe, R. T., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social networks and the performance of individuals and groups. *Academy of Management Journal*, *44*(2), 316–325.

Stanton, N. A., & Roberts, A. P. J. (2017). Examining social, information, and task networks in submarine command and control. *IEEE Transactions on Human-Machine Systems*, *48*(3), 252–265.

Stanton, N. A., & Roberts, A. P. J. (2020). Better together? Investigating new control room configurations and reduced crew size in submarine command and control. *Ergonomics*, *63*(3), 307–323. https://doi.org/10.1080/00140139.2019.1654137

Stanton, N. A., Walker, G. H., & Sorensen, L. J. (2012). It's a small world after all: Contrasting hierarchical and edge networks in a simulated intelligence analysis task. *Ergonomics*, *55*(3), 265–281. https://doi.org/10.1080/00140139.2011.642006

Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat-Rigidity Effects in Organizational Behavior: A Multilevel Analysis. *Administrative Science Quarterly*, *26*, 501–524.

Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *The American Journal of Psychology*, *77*(2), 206–219.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In *Handbook of human factors and ergonomics, 3rd ed* (pp. 243–268). John Wiley & Sons Inc. https://doi.org/10.1002/0470048204.ch9

Urban, J. M., Bowers, C. A., Monday, S. D., & Morgan Jr, B. B. (1995). Workload, Team Structure, and Communication in Team Performance. *Military Psychology*, *7*(2), 123–139. https://doi.org/10.1207/s15327876mp0702_6

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. In G. Salvendy, *Handbook of human factors and ergonomics* (4th ed., pp. 243–273). John Wiley.

Walker, G. H., Gibson, H., Stanton, N. A., Baber, C., Salmon, P., & Green, D. (2006). Event analysis of systemic teamwork (EAST): A novel integration of ergonomics methods to analyse C4i activity. *Ergonomics*, *49*(12–13), 1345–1369. https://doi.org/10.1080/00140130600612846

Walker, G. H., Stanton, N. A., Stewart, R., Jenkins, D., Wells, L., Salmon, P., & Baber, C. (2009). Using an integrated methods approach to analyse the emergent properties of military command and control. *Applied Ergonomics*, *40*(4), 636-647.

Wasserman, S., Faust, K., & Urbana-Champaign), S. (University of I. W. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering Psychology and Human Performance*. Psychology Press.

Wickens, C. D., & McCarley, J. S. (2007). *Applied Attention Theory*. CRC Press.

Wickens, C., & Tsang, P. S. (2015). Workload. In *APA handbook of human systems integration* (pp. 277–292). American Psychological Association. https://doi.org/10.1037/14528-018

Woods, D. D., & Branlat, M. (2010). Hollnagel's test: Being 'in control' of highly interdependent multi-layered networked systems. *Cognition, Technology & Work*, *12*(2), 95–101. https://doi.org/10.1007/s10111-010-0144-5

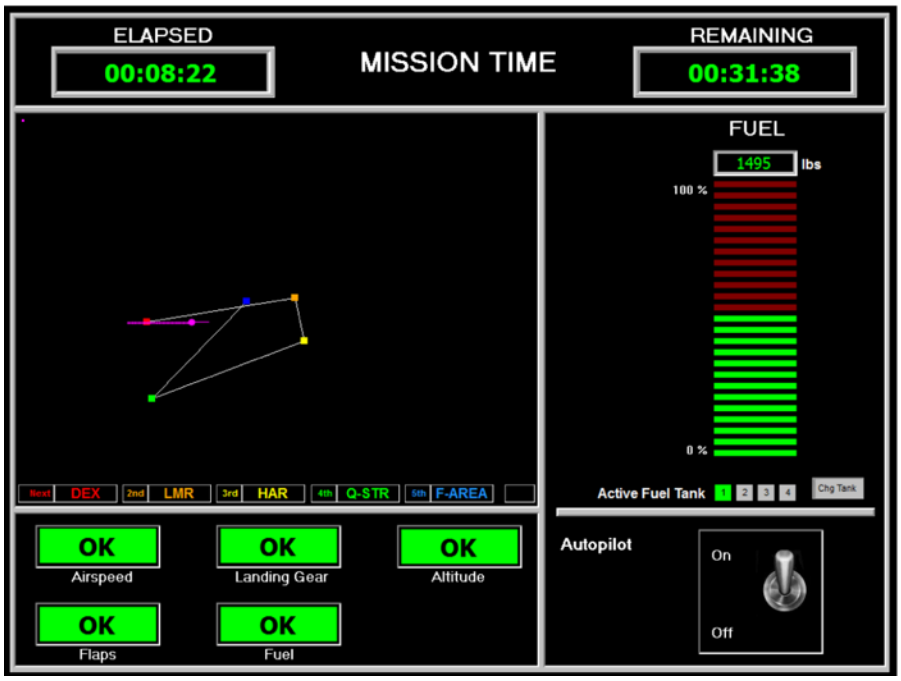
Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. https://doi.org/10.1080/00140139.2014.956151

Young, M. S., Stanton, N. A., Walker, G. H., Jenkins, D. P., & Salmon, P. M. (2007). Mental Workload in Command and Control Teams: Musings on the Outputs of EAST and WESTT. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 455–464). Springer. https://doi.org/10.1007/978-3-540-73331-7_50

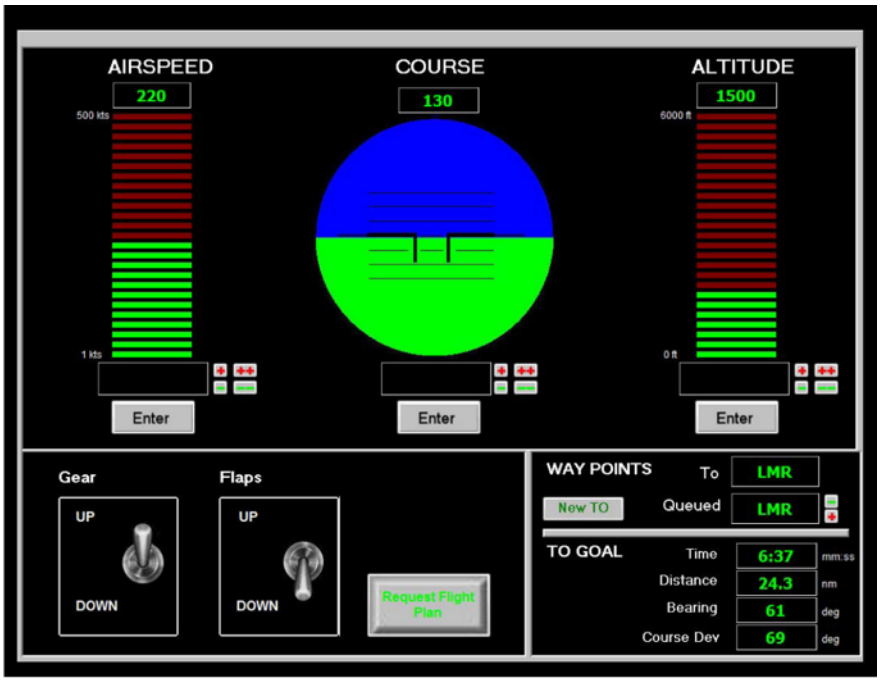
APPENDIX A

WORKSTATION INTERFACES

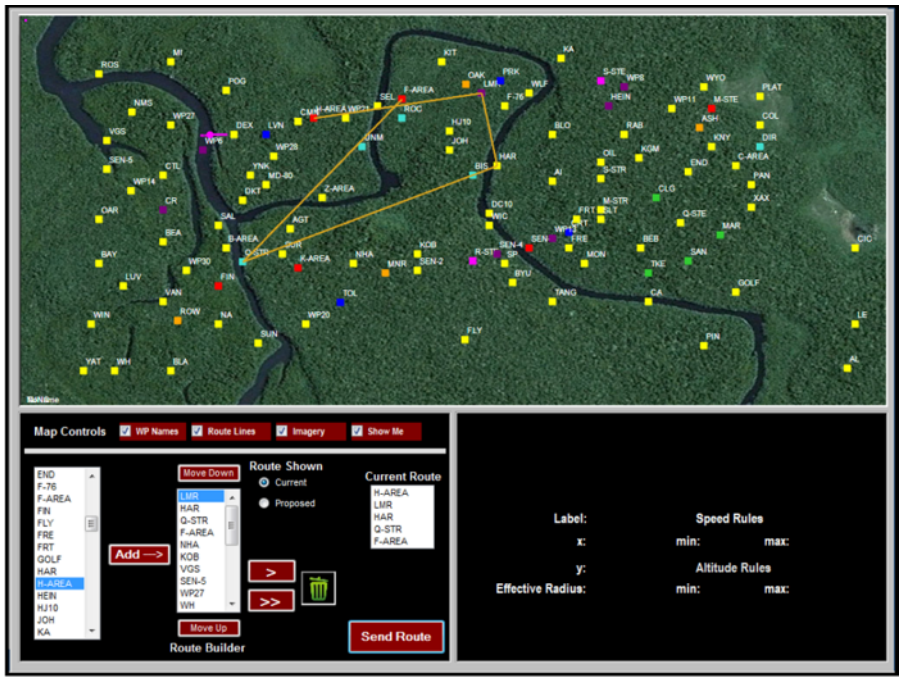
Pilot (AVO) screen 1



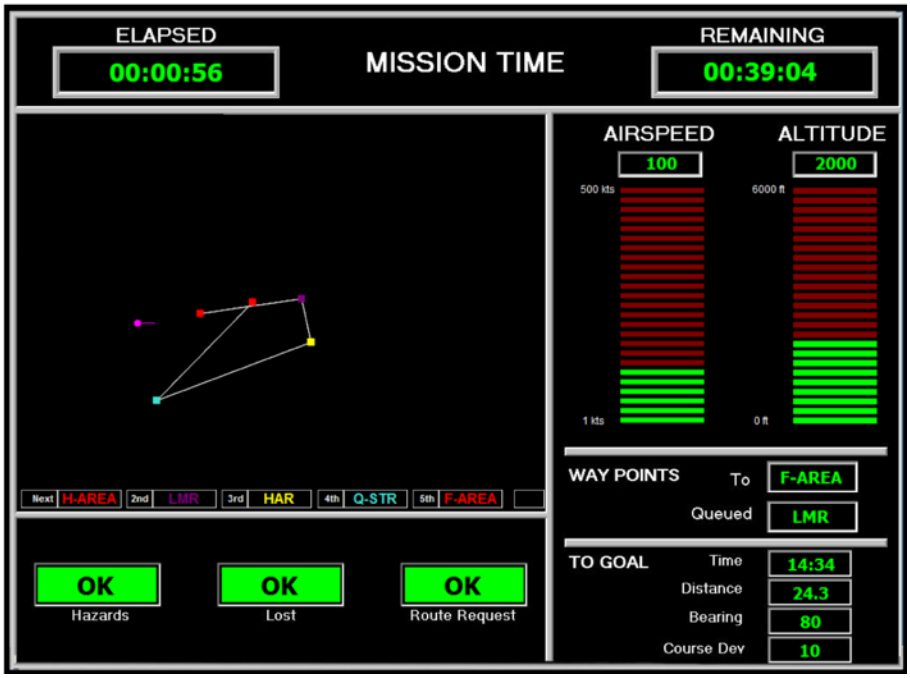
Pilot (AVO) screen 2



Navigator (DEMPC) screen 1



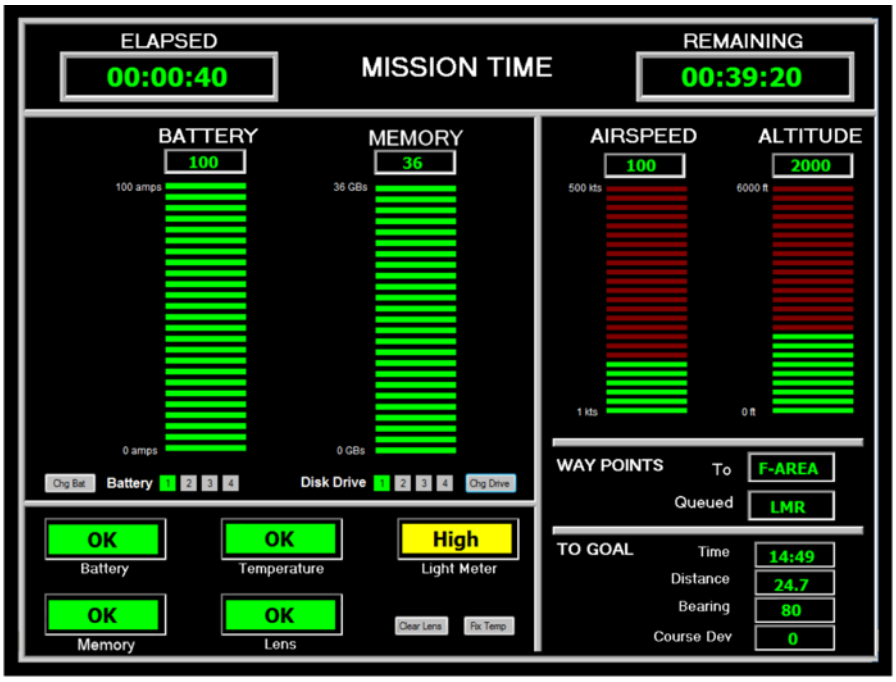
Navigator (DEMPC) screen 2



Photographer (PLO) screen 1



Photographer (PLO) screen 2



APPENDIX B

PERFORMANCE SCORE CALCULATION

The composite performance score (aka "Team Performance Score") was calculated using the following equation, where the sum of team penalties *j* are subtracted from 1000 (Cooke, Gorman, Pedersen, et al., 2007; McNeese et al., 2018).

$$Composite\ Score_i = 1000\ \left(\sum Team\ Penalty_j\right)$$

Team penalties were derived from the following four weighted penalty formulas. All times are in seconds.

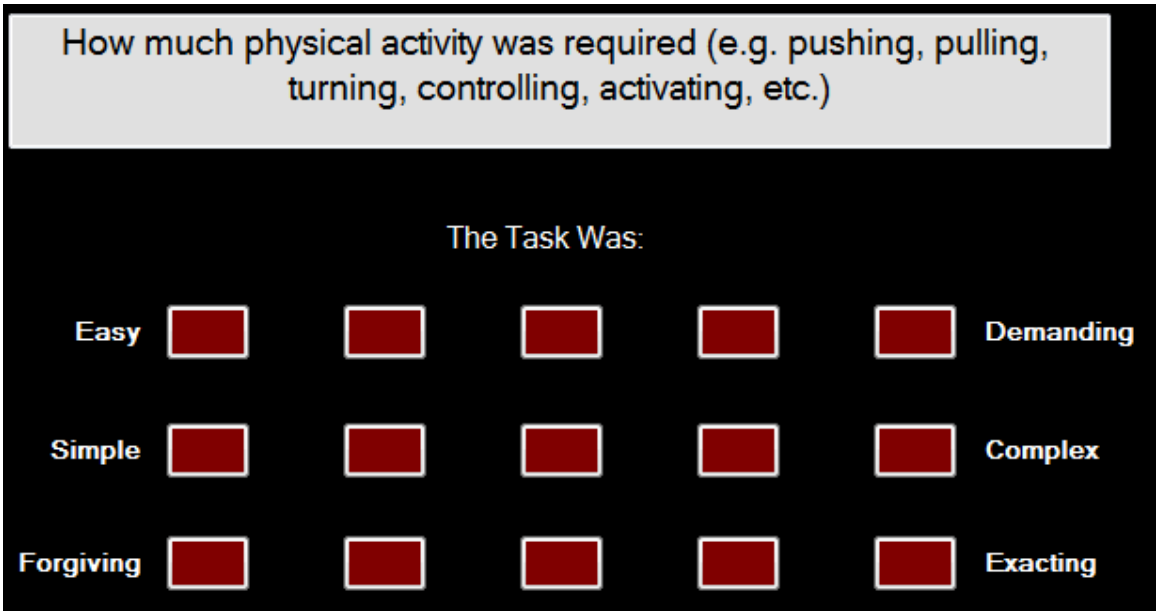
Penalty	Formula
Alarm penalty	$\left(\frac{\text{total team alarm duration}}{\text{total mission time}}\right)^{1/2} * 393.22$
Warning penalty	$\left(\frac{\text{total team warning duration}}{\text{total mission time}}\right)^{1/2} * 112.02$
Missed of slow critical waypoints penalty	$\left[1 - \frac{(\text{critical reached waypoint} * 60)}{\text{total mission time}}\right] * 318.63$
Missed or slow photos	$\left[1 - \frac{(\text{total good unique photos} * 60)}{\text{total mission time}}\right] * 314.96$

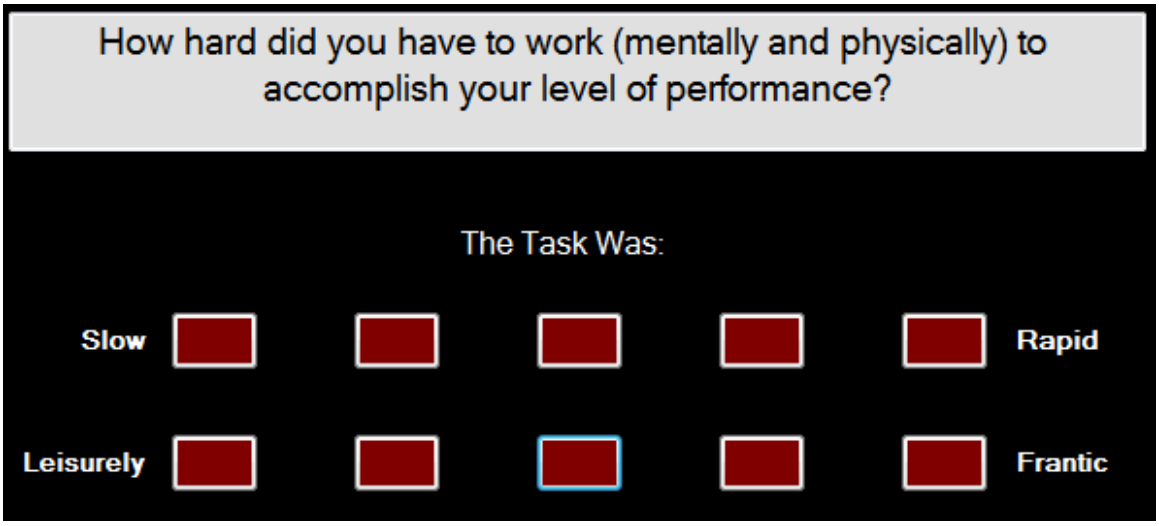
APPENDIX C

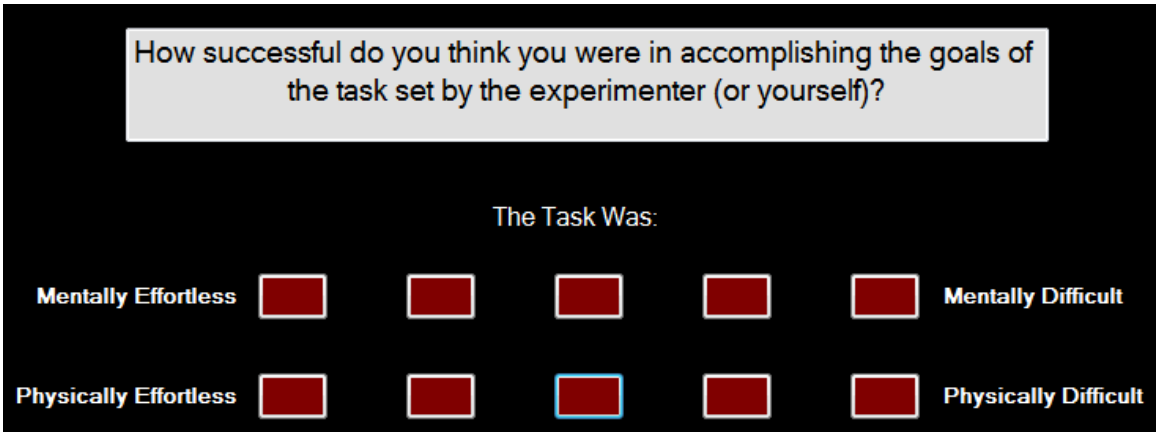
MODIFIED NASA-TLX QUESTIONAIRE

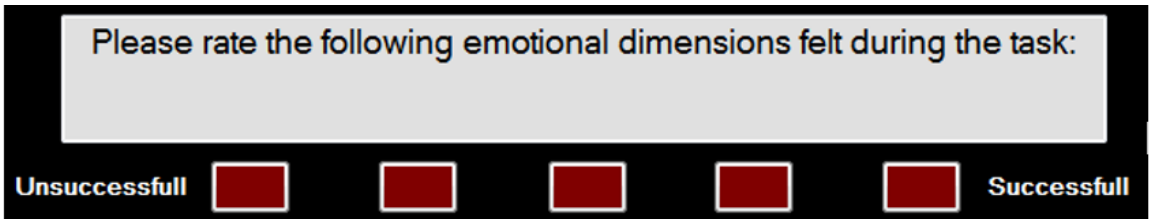
Briefing. "The evaluation you are about to perform is a technique that has been developed by NASA to assess the relative importance of six factors in determining how much workload you experienced. The procedure is simple. You will be presented with a session of pairs of rating scale titles and asked to rate which of the items was more important to your experience of workload in the task(s) that you just performed. "

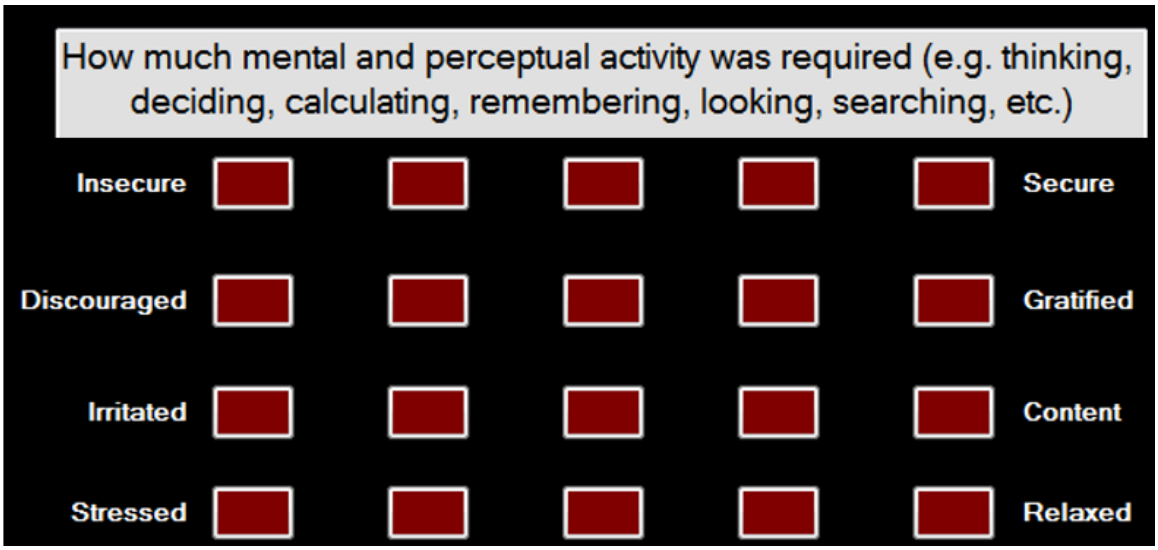
Questions. Non-numbered Likert-scale with five options presented on a touch screen interface.

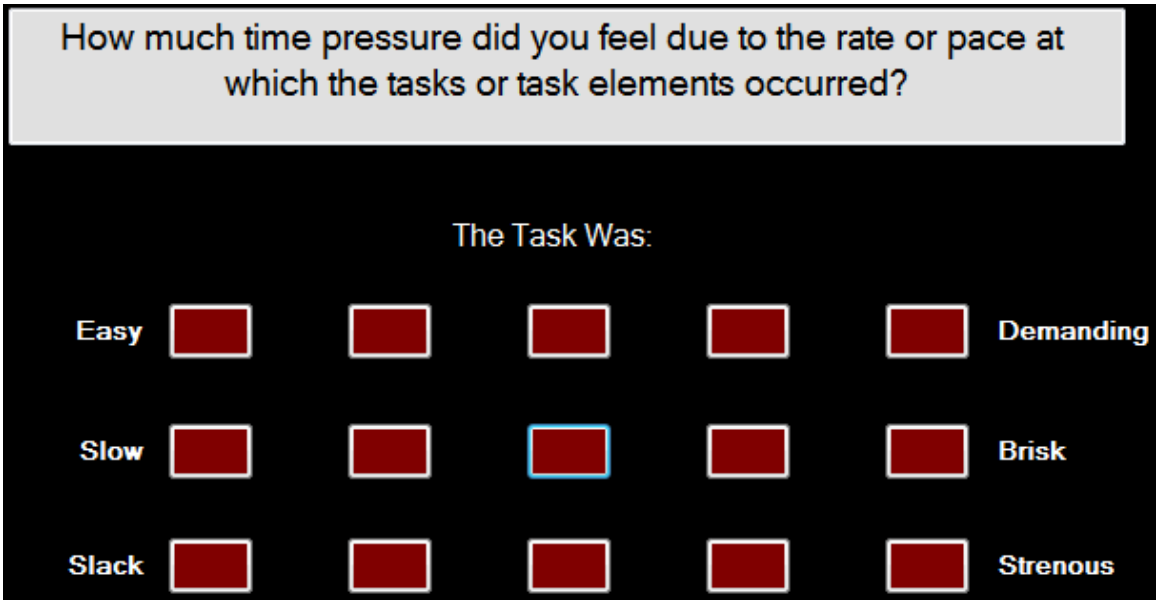





How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?
The Task Was:
Mentally Effortless
Mentally Difficult
Physically Effortless
Physically Difficult


Please rate the following emotional dimensions felt during the task:
Unsuccessfull
Successfull

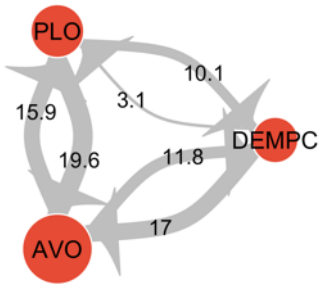
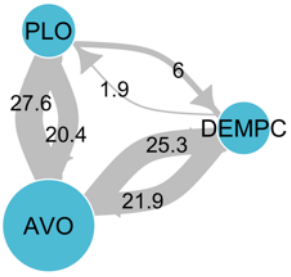
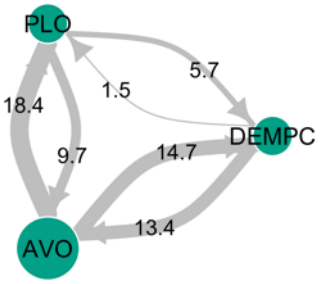
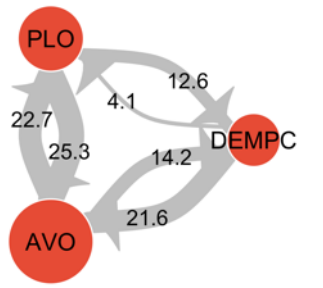
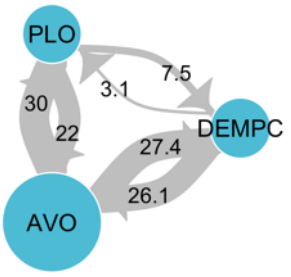
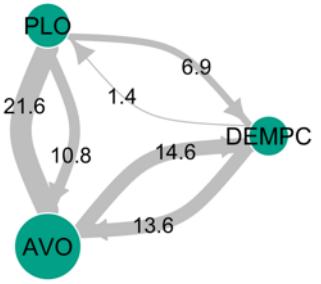

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)
Insecure
Secure
Discouraged
Gratified
Irritated
Content
Stressed
Relaxed

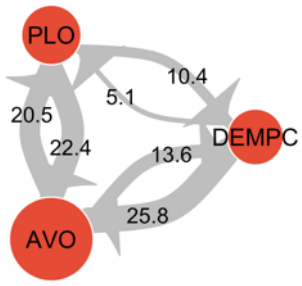
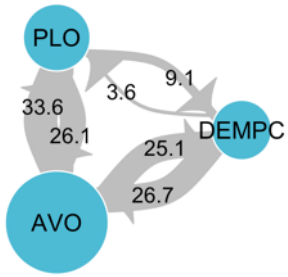
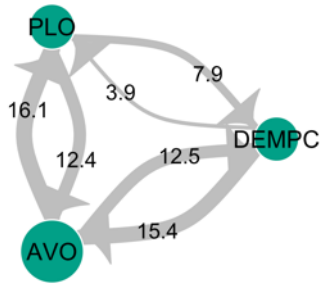
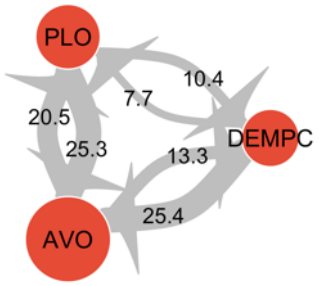
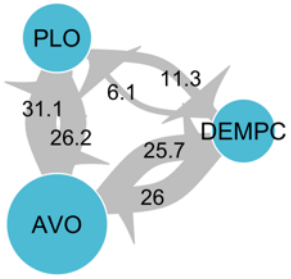
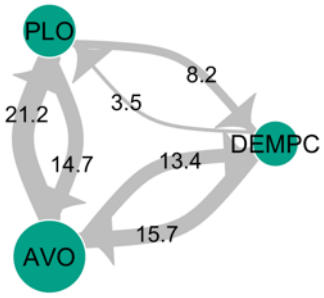

How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?
The Task Was:
Easy
Demanding
Slow
Brisk
Slack
Strenous

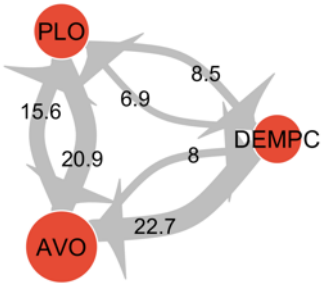
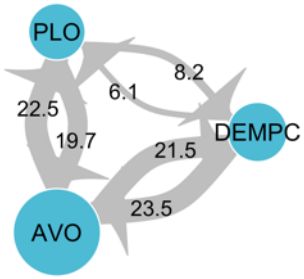
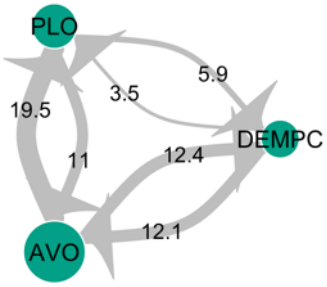
APPENDIX D

SUMMARY NETWORK DESCRIPTIONS

Summary network descriptions of conditions and missions. Each cell summarizes 10 observations. Node size indicates average weighted degree. Link width and label indicates link weight. I = intensity, C_{SD} = centrality standard deviation, C_I = indegree centralization, C_O = outdegree centralization, AVO = pilot, PLO = photographer, and DEMPC = navigator.

		Control				Experimenter				Synthetic			
Mission 1													
		Min	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**
	I	1.30	1.94	2.50	0.42	1.73	2.58	3.53	0.47	0.78	1.59	2.43	0.46
	C_{SD}	2.08	13.71	21.83	7.00	14.00	23.71	30.45	5.47	7.51	12.45	18.68	3.54
	C_I	0.09	0.23	0.40	0.12	0.07	0.12	0.23	0.06	0.02	0.10	0.32	0.08
	C_O	0.03	0.16	0.25	0.07	0.14	0.27	0.38	0.07	0.10	0.32	0.56	0.13
Mission 2													
		Min	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**
	I	1.73	2.51	5.95	1.24	2.35	2.90	3.95	0.48	1.28	1.72	2.45	0.40
	C_{SD}	3.61	18.03	41.20	10.55	17.90	2.54	4.28	0.74	6.81	13.61	20.66	4.68
	C_I	0.05	0.24	0.39	0.11	0.04	0.15	0.27	0.07	0.01	0.12	0.20	0.06
	C_O	0.09	0.20	0.41	0.11	0.11	0.25	0.42	0.09	0.04	0.31	0.45	0.12

		Control				Experimenter				Synthetic			
Mission 3													
		Min	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**
	I	1.10	2.45	4.88	1.05	2.48	3.11	4.73	0.63	0.80	1.71	3.03	0.73
	C_{SD}	9.29	17.17	47.15	11.09	6.08	23.71	30.45	5.47	2.00	11.41	22.23	6.46
	C_I	0.04	0.29	0.46	0.12	0.06	0.16	0.26	0.07	0.06	0.16	0.38	0.11
	C_O	0.03	0.16	0.41	0.15	0.04	0.23	0.44	0.12	0.06	0.23	0.49	0.12
Mission 4													
		Min	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**
	I	1.43	2.57	4.40	1.03	2.48	3.16	4.63	0.73	1.28	1.92	2.85	0.50
	C_{SD}	6.08	17.50	45.21	10.48	8.02	25.22	33.78	5.84	4.58	13.68	29.67	7.09
	C_I	0.04	0.30	0.55	0.17	0.06	0.16	0.24	0.07	0.05	0.11	0.20	0.05
	C_O	0.08	0.19	0.51	0.14	0.03	0.20	0.44	0.12	0.08	0.26	0.39	0.10

		Control				Experimenter				Synthetic			
Mission 5													
		Min	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**	**Min**	**Mean**	**Max**	**SD**
	I	0.68	2.07	3.13	0.79	1.60	2.54	4.28	0.74	0.78	1.61	2.75	0.63
	C_{SD}	5.29	12.86	23.39	6.15	6.08	25.74	37.17	9.14	7.23	12.09	16.50	3.26
	C_I	0.15	0.34	0.61	0.17	0.06	0.17	0.32	0.11	0.04	0.10	0.18	0.05
	C_O	0.04	0.19	0.32	0.11	0.03	0.18	0.41	0.13	0.02	0.27	0.47	0.12

BIOGRAPHICAL SKETCH

Craig J. Johnson graduated from Clemson University in 2013 where he obtained a Bachelor of Science in Psychology and Minor in Military Leadership. Upon graduation, he was commissioned as an active duty Armor Officer in the United States Army. While on active duty, he served as a M1A2 Abrams tank platoon leader, assistant operations officer, and company executive officer. He left active duty in 2018 and is now a Human Systems Engineering PhD student and Ira A. Fulton Schools of Engineering Dean's Fellow at Arizona State University, where he studies under Professor Nancy J. Cooke at the Center for Human, Artificial Intelligence, and Robot Teaming (CHART) and is also a member of the Cognitive Engineering Research on Team Tasks (CERTT) Lab.