Influence of Resource Allocation on Teamwork and Performance in an Intelligence, Surveillance, and Reconnaissance (ISR) "Red / Blue" Exercise Within Self-Organizing Teams

A dissertation
submitted by

James C. Won

B.S., Mechanical Engineering
Massachusetts Institute of Technology, 1997

S.M., Mechanical Engineering
Massachusetts Institute of Technology, 2001

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in

Mechanical Engineering / Human Factors Engineering

TUFTS UNIVERSITY

May 2012

ADVISER: Professor Daniel J. Hannon

This work is sponsored by the Department of the Air Force under Air Force Contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

Abstract

As task complexity and demands increase, instead of placing the entire burden on individuals, organizations are placing more emphasis on teams. This is particularly relevant in the nation's Intelligence, Surveillance, and Reconnaissance (ISR) community, where identifying clandestine networks in cultural "clutter" is one of the most complex and challenging tasks for counter-terror / counter-insurgency operations. MIT Lincoln Laboratory has developed an ISR "Red/Blue" exercise in which teams work to discover a complex network within a simulated urban environment. Teams use wide-area persistent surveillance data and decision support tools to trace relationships between individuals, events, and sites. Using this exercise, the Human Factors and Cognitive Systems Engineering Laboratory at Tufts University has investigated the influence of resources on teamwork and performance, modeling it as a two-stage decision process, and using Signal Detection Theory (SDT) as a framework to describe performance at each stage and derive metrics that describe teamwork. Team performance and teamwork are investigated within the naturalistic behavior of self-organizing teams, with different organization types, teamwork behaviors, and communication interactions that are promoted by resource allocation.

The affordances provided by available resources drive the mechanisms for communication and collaboration that distinguish the different team types. The research was executed in two phases. Phase 1 experiments involved 46 teams, of varying team size (1, 3, 4, 6, 8), and number of computers per team (1, 2, 3, 4, 6). Results from Phase 1 showed that increasing resources (people and computers) had the potential to improve performance, but once team size grew beyond an optimal size, it caused degradation in performance. Phase 1 also showed that balanced communication interactions amongst team members were indicative of better teamwork. This phase also demonstrated that the exercise, as a two-stage process, could be decomposed into taskwork and teamwork components. Phase 2 extended the study by focusing on the teamwork component of the process within 3-person teams. By holding the team size variable constant, the investigation specifically studied the effect of resource allocation (1, 2, or 3 computers) on teamwork, organization, and performance. Phase 2 results showed that providing each team member his/her own information source (computer), which provided each person the direct ability to produce and process information, resulted in increased teamwork and performance. The indication, then, in designing high performing teams, would be to facilitate each person's ability to acquire, generate, process, and share their own information as active contributors to the team process and performance.

Acknowledgements

I would first of all like to thank my advisor, Dan Hannon, who has guided me throughout the research with unwavering support, instilling my trust in “the process,” and for teaching me so much about so many things, I am truly grateful. I would also like to thank my committee members, Daniel Serfaty, Remco Chang, and Jason Rife, for all of the insights, feedback, and support that have helped make this possible, as well as my fellow labmates, Ashley and Maureen. Additional thanks to MIT Lincoln Laboratory and the Lincoln Scholars Program for providing me this opportunity, and promoting the value of human factors and cognitive systems engineering research. A special thanks to Gary Condon and Lou Bellaire for providing me with the “Red/Blue” exercise and the support that was needed to the research, and to Andy Wang for all the help with the game, and to all those who have worked to develop the game. And thanks to Lee Spence, my LSDP mentor, for being my supporter at the Lab and encouraging me to pursue this research.

Much gratitude is owed to the family of God whose prayers and support sustained me throughout. Thanks to Pastor Paul and Becky Jdsn, who have already taught me so much about teamwork for the past 19 years through His church, the family of God, as well as Pastor Dave and Angela Smn for the countless ways in which they have loved and supported me all these years, and to all my Antioch family, for all of the prayers, support, and encouragement.

I would also like to thank Pastor William, Maria Smn, and all my PMBC family in Philly, for all their love, support, and most of all, prayers that have strengthened and sustained me during all the trips up to Boston.

I would like to thank my father-in-law and mother-in-law, and all my in-laws, as well as my brother, for their prayers and support.

I would especially like to thank my parents, who have loved me and sacrificed so much for my sake all these years, and no matter what have continued to always support and encourage me through and through.

A special thanks to my kids, Ezra, Eunice, and Joanna, and a special, special thanks to my wife, Susie, for always praying for me, being my burden-sharer, and taking on so much to help and encourage me through everything.

Lastly, thank you Jesus, for allowing me to embark on this journey, for showing me how to lean upon your strength, follow in your steps, and trust you in all things.

And we know that in all things God works for the good of those who love him, who have been called according to his purpose. (Romans 8:28)

Table of Contents

Acknowledgements vi

List of Figures xiii

List of Tables xvii

1 INTRODUCTION

1.1 Team Performance and Teamwork 4

1.2 Teamwork 6

1.2.1 Communication and Information Sharing 9

1.2.2 Memory 15

1.2.3 Leadership 19

1.2.4 Team Structure 21

1.2.5 Self-Organizing Teams 24

1.2.6 Team Size 26

1.3 Performance Metrics 30

1.3.1 Situational Awareness Measure: SART 30

1.3.2 Workload Measure: NASA-TLX 31

1.4 Workload 33

1.4.1 Workload and Performance 34

1.4.2 Workload and Situational Awareness 37

1.4.3 Team Workload 45

1.5 ISR "Red/Blue" Exercise 50

1.6 Information Theory 54

1.7 Signal Detection Theory 55

1.8 Research Objectives 60

2 GENERAL METHOD

2.1 Participants 64

2.2 Apparatus 65

2.3 Procedure 70

2.3.1 Clandestine Network Construction Workflow 71

2.3.2 Detailed Task Description 74

2.3.3 Two-stage decision process 85

2.4 Performance Metrics 87

3 PHASE 1

3.1 Methodology 92

3.1.1 Participants 92

3.1.2 Procedure 94

3.2 Results 94

3.2.1 Detection Stage, Decision Stage, Overall Sensitivity (*d'*) 100

3.2.2 Delta decision (Δ_{dec}) and Delta total (Δ_{tot}) 103

3.2.2.1 Team Categories 105

3.2.3 Impact of Resources 111

3.2.4 Communication Interactions 114

3.2.5 Subjective Assessments 116

3.3 Discussion 120

3.3.1 Information Production/Information Synthesis 126

3.3.2 Conclusions 127

4 PHASE 2

4.1 Methodology 133

4.1.1 Participants 133

4.1.2 Procedure 135

4.1.3 Variables ... 135
4.1.4 Research Questions ... 138
4.2 Results ... 147
4.2.1 One-Computer Teams ... 151
4.2.2 Two-Computer Teams ... 154
4.2.3 Three-Computer Teams ... 157
4.2.4 Impact of Resources ... 159
4.2.4.1 Delta decision (Δ_{dec}) and Delta total (Δ_{tot}) ... 163
4.2.5 Communication Interactions ... 166
4.2.6 NASA-TLX and SART ... 170
4.3 Discussion ... 176
5 DISCUSSION AND CONCLUSION
5.1 Closed-loop system ... 185
5.2 Delta ... 187
5.3 Effect of Resources ... 188
5.4 Practical Implications ... 192
5.5 Future Work ... 193
5.6 Conclusion ... 199
6 APPENDICES
Appendix A: Individual Participant (1p) Descriptives ... 205
Appendix B: 4-Person Team (4p) Descriptives ... 206
Appendix C: 6-Person Team (6p) Descriptives ... 207
Appendix D: 8-Person Team (8p) Descriptives ... 208
Appendix E: 1-Computer Team (1C) Descriptives ... 209
Appendix F: 2-Computer Team (2C) Descriptives ... 210
Appendix G: 3-Computer Team (3C) Descriptives ... 211

Appendix H: Results by A' 212
Appendix I: Command, Task Force Blue Latern 214
Appendix J: Scoring Matrix 215
Appendix K: NASA-TLX and SART Questionnaires 216
Appendix L: Team Workload, Teamwork, and Mutual Helpfulness Questionnaires 217
Appendix M: Training to Criteria 221
References 225

List of Figures

Figure 1.1 *Workload-Performance curve (Cummings & Nehme, 2010)* .. 35

Figure 1.2 *Diagram of Perceptual Control Theory (Hendy, 1995)* 40

Figure 1.3 *ISR "Red/Blue" exercise software tool: "BlueStreak"* 53

Figure 1.4 *Four outcomes of Signal Detection Theory* 56

Figure 1.5 *Signal and noise distributions (Green & Swets, 1966)* 57

Figure 2.1 *Experiment room setup* 66

Figure 2.2 *"BlueStreak" map view* 67

Figure 2.3 *"Bluestreak" graph view* 68

Figure 2.4 *Diagram for a typical network construction workflow* 72

Figure 2.5 *Example of breadth-first search* 73

Figure 2.6 *Example of depth-first search* 74

Figure 2.7 *Initial map view* 75

Figure 2.8 *Message panel* 76

Figure 2.9 *Message details* 77

Figure 2.10 *Creating a "placemark" at location of interest* 78

Figure 2.11 *"New Placemark" window* 79

Figure 2.12 *"Create nomination" feature* 81

Figure 2.13 *Nomination window* 82

Figure 2.14 *Nominated tracks* 83

Figure 2.15 *Subsequent placemarks and nominations* 84

Figure 2.16 *Two stage decision process* 85

Figure 3.1 *Phase 1 results on ROC graph* 99

Figure 3.2 *Performance (d' values) by team size (mean values)* 101

Figure 3.3 *Delta decision (Δ_{dec}) and Delta total (Δ_{tot}) by team size (mean values)* 104
Figure 3.4 *Delta decision (Δ_{dec}) by team size* 106
Figure 3.5 *Delta total (Δ_{tot}) by team size* 106
Figure 3.6 *d' values by team categories (mean values)* 110
Figure 3.7 *Number of Hits (H) or False Alarms (FA) by team size (mean values)* 111
Figure 3.8 *Number of Hits (H) or False Alarms (FA) by team size (mean values)* 113
Figure 3.9 *Number of Hits (H) or False Alarms (FA) by team category (mean values)* 114
Figure 3.10 *Communication Interaction Ratio (CIR)* 115
Figure 3.11 *Communication Interaction Ratio (CIR) by Delta decision (Δ_{dec})* 116
Figure 3.12 *NASA-TLX ratings by team size (mean values)* 117
Figure 3.13 *SART ratings by team size (mean values)* 119
Figure 3.14 *Correlation between Delta total (Δ_{tot}) and Delta decision (Δ_{dec}) in Phase 1* 124
Figure 4.1 *Three team organization structures via resource allocation*..139
Figure 4.2 *Phase 2 results on ROC graph* 148
Figure 4.3 *1-Computer (1C) team results (d' values, mean)* 151
Figure 4.4 *1-Computer (1C) team results (d' values) identified by Δ_{tot}* ..152
Figure 4.5 *Normalized 1-Computer (1C) team results (d' values) identified by Δ_{tot}* 153
Figure 4.6 *2-Computer (2C) team results (d' values, mean)* 154
Figure 4.7 *2-Computer (2C) team results (d' values) identified by Δ_{tot}* ..155

Figure 4.8 *Normalized 2-Computer (2C) team results (d' values) identified by* Δ_{tot} 156
Figure 4.9 *3-Computer (3C) team results (d' values, mean)* 157
Figure 4.10 *3-Computer (3C) team results (d' values) identified by* Δ_{tot} 158
Figure 4.11 *Normalized 3-Computer (3C) team results (d' values) identified by* Δ_{tot} 159
Figure 4.12 *Phase 2 results (d' values) categorized by resource allocation (mean)* 160
Figure 4.13 *Phase 2 results (d' values) categorized by resource allocation (mean)* 161
Figure 4.14 *Normalized Phase 2 results (d' values) categorized by resource allocation (mean)* 162
Figure 4.15 *Delta decision (Δ_{dec}) and Delta total (Δ_{tot}) by resource allocation* 163
Figure 4.16 *Delta decision (Δ_{dec}) by resource allocation* 164
Figure 4.17 *Delta total (Δ_{tot}) by resource allocation* 165
Figure 4.18 *Correlation between Delta total (Δ_{tot}) and Delta decision (Δ_{dec}) in Phase 2* 166
Figure 4.19 *Communication Interaction Ratio (CIR) by Delta decision (Δ_{dec}) (mean)* 167
Figure 4.20 *Communication Interaction Ratio (CIR) by resource allocation* 169
Figure 4.21 *NASA-TLX ratings by resource allocation (mean values)* 170
Figure 4.22 *NASA-TLX ratings by resource allocation (all teams)* 172

Figure 4.23 *NASA-TLX ratings by resource allocation, with individual participant results from Phase 1 (all teams)*....................174

Figure 4.24 *SART ratings by resource allocation (mean values)* 175

Figure 4.25 *Normalized Phase 2 results (d' values) by resource allocation with Phase 1 individual participant results (mean)*..........179

Figure 5.1 *Simple closed-loop control system* 185

Figure 5.2 *Team performance and teamwork as a closed-loop control system*...186

Figure 5.3 *Normalized Phase 2 results (d' values) by resource allocation with Phase 1 individual participants and 6-person team results (mean values)*..191

List of Tables

Table 1.1 *Measures of Verbal Communication*..................................10

Table 1.2 *NASA-TLX definitions (Hart & Staveland, 1988)* 33

Table 3.1 *Phase 1 summary* 93

Table 3.2 *Phase 1 results summary (mean values)* 95

Table 3.3 *Phase 1 results by team categories (mean values)*...........108

Table 4.1 *Phase 2 summary* 134

Table 4.2 *Independent Variable (IV) and Dependent Variables (DV)* 138

Table 4.3 *Potential outcomes of Phase 2* 142

Table 4.4 *Phase 2 results summary (mean values)* 147

1

Introduction

With the advancement of technology and complexity of tasks which organizations encounter, the difficulty of decision making and execution of tasks has correspondingly increased. Rather than placing this entire burden on individuals, it has caused a shift in emphasis to teams and to the importance of performance in teams (Mesmer-Magnus & DeChurch, 2009), with a focus on understanding effective coordination and collaboration that facilitates this performance (Cannon-Bowers & Salas, 1990). Hence, from the early 1980's to the present, the use of teams in organizations has dramatically increased (Hollenbeck, DeRue & Guzzo, 2004).

The word “team” is used, in order to distinguish it from the word “group”. The difference is subtle, but Orasanu and Salas (1993) define “group” as a simple collection of people, who may share something in common (Massey, 1991). On the other hand, a “team” can be seen as a special class of groups (O’Neil, Baker, & Kazlauskas 1992). Teams can be seen as groups that are goal-oriented, and share a common goal among team members (Ilgen, Major, Hollenbeck & Sego 1995; Johnson, Suriya, Yoon, Berrett & Fleur 2002; Orasanu & Salas 1993).

A “team” has been defined as “a distinguishable set of two or more individuals who interact interdependently and adaptively to achieve specified, shared, and valued objectives” (McIntyre & Salas, 1995; Salas, Stagl, Burke, & Goodwin, 2007). Intelligence analyst cells, air traffic control flight crews, hospital surgery teams, sports teams, church planting teams (Kim & Martin, 2009), and classroom working groups are some examples of the many types of teams.

Teams are being utilized more and are seen as beneficial and useful because they are especially skilled in operations in dynamic and complex situations (Burke, Stagl, Salas, Pierce, & Kendall, 2006). In those kinds of situations, a team versus an individual's ability to rapidly adapt and respond gives it an advantage in performance.

These advantages make the emphasis on teams particularly applicable to the nation's Intelligence, Surveillance, and Reconnaissance (ISR) community, where identifying clandestine networks in cultural "clutter" is one of the most challenging tasks for counter-terror/counter-insurgency operations (Won, Condon, Landon, Wang, & Hannon, 2011). These types of teams perform time sensitive tasks in dynamic environments (Burke, Salas, Wilson-Donnelly, & Priest, 2005) by synchronizing their efforts to meet team level goals (Cannon-Bowers, Salas, & Converse, 1993; Cooke, Gorman, Duran, & Taylor, 2007) and build a common team-level tactical picture (Tollner, 2009). ISR technologies such as persistent surveillance assets can provide the raw data needed to support this task of network discovery. However, advanced technologies and algorithms are not able to fully automate the exploitation and decision making process, which is the whole process of producing data, synthesizing data, processing and making sense of it, and ultimately, making decisions based upon it. Thus, it is recognized that this complex network discovery process is inherently human-centric, and with the large amounts of data that are involved, inevitably relies upon teams of intelligence analysts working together towards this goal.

The mission area of ISR was a target of opportunity for the present research, and provided an operational platform to perform team research. Team research itself is a complex field, with many different threads, and requires the bringing together of many disparate notions. And even though there may be some disconnect between certain elements, the overarching goal of this research is to gain a deeper overall understanding of how teams perform from a human factors perspective; effectively, how they utilize resources and information in team organization, teamwork, and overall performance.

1.1 Team Performance and Teamwork

Team performance can be seen as a product of the interactions and coordination amongst the members of a team, and the dynamics of these processes create an emergent property that is greater than the sum of the individual attributes (Bandura, 1997).

The power that teams offer is demonstrated in a variety of situations, where individuals are able to come together and perform at a synergistic level that is higher than a simple aggregation of each individual's performance. Such observations are plentiful in sports, where terms such as "team chemistry" exist, in which a team of players with team chemistry, but with no clear "superstar," is able to defeat a disjointed team with one superstar who does all of the work.

Woolley, Chabris, Pentland, Hashmi, and Malone (2010) demonstrated a concept called "collective intelligence" that exists in teams of people working together. In their study, teams worked together on tasks that ranged from visual puzzles to negotiations, brainstorming, games and complex rule-based design assignments. This "collective intelligence" was found to be a property of the team itself, not just the individuals in it, and was found to be an excellent predictor of team performance in complex tasks (Woolley et al., 2010).

Likewise, in regard to memory performance, research has indicated that team memory performance is superior to individual memory. Hinsz (1990) proposed three reasons for this: (1) teams have a larger pool of available information because groups can rely on each other's memory, (2) teams correct the errors of their members, whereas there is no such process with individuals, and (3) teams have more effective decision making processes on judgments regarding remembered information. In the ISR simulation exercise for this research, such memory effects may underlie various team member behaviors, as well as the team's collective decision making processes.

In recent years, many models and theories have been developed in an attempt to understand the causes of effective performance in higher performing teams, to uncover the underlying sources of performance outcomes and provide improvements to team processes (Salas et al., 2007). And while there has been no definitive, unanimous consensus, in a widely accepted study, Salas, Sims, & Burke (2005), through their review of team literature over the past twenty years, have stated *teamwork* as the key factor in team performance, and have developed a model of teamwork based upon their review.

1.2 Teamwork

Salas et al. (2005) have defined teamwork as "a set of interrelated thoughts, actions, and feelings of each team member that are needed to function as a team and that combine to facilitate coordinated, adaptive performance and task objectives resulting in value-added outcomes."

They have identified five core components of teamwork: team leadership, mutual performance monitoring, backup behavior, adaptability, and team orientation. They have also identified three coordinating mechanisms that facilitate teamwork: shared mental models, closed-loop communication, and mutual trust (Salas et al., 2005). As teamwork has been found to be the key to team performance, Salas et al. (2005) state that teams which employ the five components of teamwork, while making sure to utilize the three coordinating mechanisms, should outperform those teams that do not.

These components of teamwork give rise to emergent concepts such as “team cognition.” The concept of team cognition has, in similar ways, captured the ideas of coordinating mechanisms of teamwork, and has been found to be a key component to successfully achieving a team’s goals (Salas, Rosen, Burke, Nicholson, & Howse, 2007). Team cognition is explained as the interaction of and dependencies between intra-individual and inter-individual level processes (Fiore & Schooler, 2004). Fiore and Salas (2004) propose that the manifestation of team cognition is the “seamless execution of coordinated behaviors,” a type of “binding mechanism”, that fuses multiple relevant information into a functional entity (Von der Malsburg, 1995). Essentially, team cognition may be the mechanism that binds together the many inputs of a team into its own functional entity (Fiore & Salas, 2004), promoting the coordination that aids problem solving (Laughlin, Hatch, Silver, & Boh, 2006)

In understanding team cognition as a binding mechanism that produces coordinated behavior, the components that comprise team cognition then become important factors in understanding teamwork. Two components, which contain particular relevance to the current research, are presented by Fiore and Salas (2004): shared awareness and team communication.

Shared awareness has to do with perceptions of "who knows what." And team communication is the process through with effective team knowledge is attained (Fiore & Salas, 2004), and therefore communication protocols can be used as a window to team cognition (Cooke, Salas, Kiekel, & Bell, 2004). In this sense, and with this perspective on teamwork, communication can be viewed as a central mechanism of information sharing and information processing (Salas, Cooke, & Rosen, 2008).

Within a team decision making process, conditions that best promote shared awareness and effective communication could be expected to yield better teamwork and performance. And these two components of team cognition are indeed observed in this research.

1.2.1 Communication and Information Sharing

Effective communication is an intuitively important concept for teamwork and team performance. Tollner (2009) showed the power of teamwork in her findings that teams have an advantage over individuals in completing a change detection task (e.g., detecting the appearance of a hostile aircraft on one's radar screen), but this advantage was only realized when teams communicated.

Some studies have extended this investigation of communication by examining the quality of communication. Their premise was that simply measuring quantity of communication was not a sufficient indication of effective team performance. Because in some cases, reduction in communication could indicate more efficient communication, their position was that true measurement of team performance needs to involve investigation of the types of communication in which teams are engaged.

To this extent, Entin and Entin (2001) proposed some common measures of verbal communication, shown in Table 1.1.

Table 1.1 *Measures of Verbal Communication*

Measure	**Description**
Overall Rate	
Total Communications	Total number of communications per minute
Communication Types	
Information Requests	Number of requests for information per minute
Information Transfers	Number of transmissions of information per minute
Action Requests	Number of requests for an action per minute
Action Transfers	Number of statements of actions (to be) taken per minute
Coordination Requests	Number of requests to coordinate an action per minute
Coordination Transfers	Number of agreements to coordinate an action per minute
Acknowledgements	Number of non-substantive acknowledgements of receipt of communication (e.g., 'ok' to acknowledge receipt of information) per minute
Communication Ratios	
Overall anticipation	All communication transfers divided by all communication requests
Information anticipation	Information transfers divided by information requests
Action anticipation	Action transfers divided by actions requests

Table 1.1 shows some common communication measures, which can be measured at both the individual or team level. The names of the measures and their descriptions are given in Table 1.1. One particular measure, the anticipation ratio, has proven to be a useful metric (Entin, Serfaty, & Deckert, 1994; Entin & Serfaty, 1999; Serfaty, Entin, & Johnston, 1998). It is the ratio of information transfers to information requests. Ratios greater than 1.0 could indicate that team members are anticipating the information request from a teammate and pushing the information to them before it is requested (Entin et al., 1994). These types of transfers can even be observed in sports teams, where actions such as "blind passes"

occur in anticipation of a team member being in a certain position, and seem to indicate good teamwork.

Others have looked at the role or purpose of communication. A more extended definition of "team" is "a distinguishable set of two or more people who interact, dynamically, interdependently, and adaptively toward a common goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life-span of membership" (Salas, Dickenson, Converse, & Tannenbaum, 1992). This definition has particular relevance with regard to information sharing. It suggests that communication interactions become the crucial element to achieving the common goal because the dynamic interaction between the different roles primarily takes place through information sharing (Schraagen, Veld, & De Koning, 2010).

The current research does not investigate types of communications, but attempts to examine communication from the perspective of its role in information sharing. In some studies, communication specifically identified as information sharing was found to impact team decision making (Bunderson & Sutcliffe, 2002; Jehn & Shah, 1997). In addition, as mentioned at the beginning of section 1.2, communication is a key component of team cognition when it serves the function of information sharing and processing. (Salas et al., 2008).

Mesmer-Magnus and DeChurch (2009) performed a meta-analysis on the topic of team information sharing (IS), which is defined as a "central process through which team members collectively utilize their available informational resources" (Mesmer-Magnus & DeChurch, 2009). Their findings revealed the importance of information sharing to team cohesion, decision satisfaction, knowledge integration, and most relevantly, team performance.

Stasser and Titus's (1985, 1987) biased information sampling model showed that groups spend more time discussing shared information than unshared information, which related to the concepts of uniqueness and openness in teams. Uniqueness captures how much teams are utilizing each member's unique knowledge for the team's benefit, whereas openness deals with team communication independent of the initial distribution of information amongst the team members (Henry, 1995; Jehn & Shah, 1997).

To test a variety of hypotheses related to IS and its impact on team performance and processes, Mesmer-Magmus and DeChurch (2009) investigated 72 independent studies as part of their meta-analysis. Some of their findings are as follows: IS positively predicted team performance, cohesion, and knowledge integration; IS uniqueness was more predictive of team performance than IS openness; IS was greater in teams where member similarity is high; and as team members became more informationally independent, more information was shared.

Their findings suggest a knowledge redundancy effect. Popular thinking regarding team makeup is that having the least knowledge-redundancy in teams, having more team diversity, is desired. However, the findings of Mesmer-Magmus and DeChurch (2009) indicate that less knowledge-redundant teams (those who can gain the most from information sharing) actually share less information than the more knowledge-redundant teams.

Information sharing has been shown in a variety of studies to have a positive impact on team performance, and an important factor in teams developing shared cognition. As described earlier, the concept of team cognition emerges through a variety of team behaviors, communication and IS being part of that emergence.

In particular, Pearsall and Ellis (2010) investigated a specific communication pattern, labeled "role identification behaviors," and its influence on team performance. Specifically, it examines the influence of role identification behaviors on team mental model and transactive memory development as they relate to team performance.

For this investigation, Pearsall and Ellis (2010) used a modified version of Aptima, Inc.'s Distributed Dynamic Decision-Making (DDD) simulation with 240 college students, divided up into 60 four-person teams. Role identification behavior was measured through coding of verbal communication regarding team members' attempts to resolve two questions: "Who requires inputs from me?" and "From whom do I require inputs?" The team-interaction mental model was measured by asking each member to fill in concepts for a team concept map that best represented actions of each team member. Transactive memory was measured on a scale developed by Lewis (2003), which has 15 items (3 dimensions), designed to assess specialization, credibility, and coordination. Team performance was measured by the team's task outcome after the simulation. The study identified role identification as a specific behavioral process that assists in the development of team cognition components, and ultimately impacts team performance. The results indicated that the specific communication identified as role identification behavior predicted the development of the team-interaction mental models and transactive memory, which impacted performance.

Ultimately, communication and information sharing have been found to be important aspects of teamwork and performance. And although they seem intuitively important to team outcomes, as in certain cases mentioned above, teams were found to not share information when they most needed to do so (Mesmer-Magnus & DeChurch, 2009). If this type of information sharing behavior is a natural tendency of team members' behavior, results such as these could point to the importance of having common access and sharing of data and information across the team.

1.2.2 Memory

As previously mentioned, memory in teams, such as transactive memory and collective memory, influences the team decision making process. Transactive memory, which represents a shared system for encoding, storing, and retrieving information (Moreland, 1999; Wegner, 1987), is essentially a metaknowledge of knowing what team members know about one another's knowledge or expertise (Solansky, 2008).

Research has indicated that a team's collective memory performance is superior to individual memory. Hinsz (1990) proposed three reasons for this: (1) teams have a larger pool of available information because groups can rely on each other's memory, (2) teams correct the errors of their members, whereas there is no such process with individuals, and (3) teams have more effective decision making processes on judgments regarding remembered information.

Hinsz (1990) applied signal detection theory (SDT) and social decision scheme (SDS) theory to examine group memory performance. For the signal detection theory aspect of recognition memory, the parameter *d'* was used to indicate a participant's ability to discriminate between pieces of information that were or were not presented, thereby giving an indication of their memory. Another parameter, *beta (β)*, was used to reflect the decision strategy of the participant. These aspects of SDT will be furthered discussed later in the chapter. For the social decision scheme theory aspect, the primary objective was to uncover the characteristics of the group decision process that combines the various positions of group members into a single group consensus position. One possible decision scheme that could be relevant is the "truth-wins" decision scheme, which predicts that if one of the group members is able to identify the correct choice, the group will be able to make the correct choice.

Hinsz (1990) used an experiment in which 6-person teams were compared with individuals in their recognition memory for material from video-taped job interviews. The findings showed that groups were more sensitive than individuals in memory for the presented information, and groups had fewer errors than individuals. However, groups and individuals did not differ in their relative proportion of miss and false alarm responses, indicating that groups make the same pattern of recognition errors as individuals, just fewer of them. Groups were also found to have higher confidence in their recognition memory than individuals.

Providing further investigation into the team advantage in recognition memory, Clark, Hori, Putnam, and Martin (2000) investigated the impact of team collaboration (versus simply pooling of information) in establishing this advantage. Their experiments varied team size to examine the role of majority processes in team recognition memory decisions. Teams of two had no majority, while teams of three always had a majority. Participants were presented with four lists, each with 24 words. An interesting finding was that group effects differed for targets and distractors. Collaboration assisted in recognizing targets but not for rejecting distractors.

While simple tasks may not require much memory load, complex tasks may require higher amounts of memory load, and this load may need to be distributed within a team. When teams are better able to coordinate this memory load distribution, they can be expected to perform better than individuals. This would be consistent with the conclusions proposed by Hinsz (1990), that teams have a larger pool of available information, they can offer corrective feedback, and they can have more effective decision making processes on judgments regarding remembered information.

It can be seen how a team's collective memory can assist in developing shared mental models, a coordinating mechanism for teamwork (Salas et al., 2005). Shared mental models are "organized knowledge shared by team members" (Orasanu & Salas, 1993), and have been found to be an important component of effective team performance (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). Within the current research, conditions that lessen the load on a team's collective memory and promote shared mental models are indeed shown to positively affect teamwork and performance.

1.2.3 Leadership

Another teamwork concept is leadership, which can be an important factor in achieving a team's goals. A team leader can influence the team's cognitive, motivational, and affective processes (Zaccaro, Rittman, & Marks, 2001) and can influence effective performance (Entin & Serfaty, 1999; Marks, Zaccaro, & Mathieu, 2000) Additionally, leadership style has been found to have a significant effect on team performance. Some research has shown the benefits of shared leadership versus a single leader in self-managed teams (Solansky, 2008), with some research suggesting that shared leadership can be an important predictor of team effectiveness (Ensley, Pearson, & Pearce, 2003).

Solansky (2008) explored leadership style and its influence on team processes, in particular, as it pertains to self-managed teams. Typically, self-managed teams do not have an authority-designated leader, but rather the team designates its own leader. Solansky (2008) specifically explored leadership processes (shared vs. unshared), and found that teams with shared leadership have higher collective efficacy (performance confidence), lower relational conflict, and higher levels of transactive memory. However, there have been other findings which demonstrate that strong, directive leaders can sometimes enhance performance (Peterson, Owen, Tetlock, Fan, & Martorana, 1998)

Other studies have further explored the concept of leadership style and joint decision making. A leader's ability to differentially utilize the recommendations of staff members was found to be important to the accuracy of team decision making (Humphrey, Hollenbeck, Meyer, & Ilgen, 2002; Hollenbeck et al., 2004; Phillips, 1999). Phillips (2002) found that allowing staff members greater process control by giving them the opportunity to influence how the leader uses their recommendations when making the final decision increased staff members' perceptions of fairness, and ultimately, the members' output.

This is but a small sampling of leadership research, but as these examples show, the effect of leadership in and of itself has many disparate theories and concepts. As much as some studies show the importance of shared leadership (Solansky, 2008; Ensley et al., 2003), there exists other studies that show the benefits of strong, directive leadership (Peterson et al., 1998). However, as leadership research progresses, it does remain to be seen, whether in many situations, the leadership effect could simply be a property that emerges from how a team is organized or structured (Shaw, 1960; Taggar & Brown, 2001). In the current research, the role of a leader is not explicitly assigned. The roles and organization of these teams are naturally forming within the structure afforded to them via the given resources. Within these naturalistic structures, leadership style is not a main focus of investigation, but the leadership effect is viewed more as an emergent property of certain team structures that develops from available resources.

1.2.4 Team Structure

Team structure is an aspect of teams and teamwork that can have a significant impact on performance. Two broadly accepted categories of team structure are *hierarchical* and *heterarchical*.

A classic hierarchy can be defined as "a structure in which members pass on information to a leader, but not to each other" (Schraagen et al., 2010), or one in which the leader is the only individual with an overall view (Walker, Stanton, Salmon, Jenkins, Monnan, & Handy, 2008). The approach of a hierarchical structure is characterized by centralized planning, decomposition of tasks, and control processes (Levchuk, Yu, Levchuk, & Pattipati, 2004). In many cases, the responsibility, authority, and discretion for the final decision reside at the top of the hierarchy (Kang, 2010; Schraagen et al., 2010).

Some examples where hierarchical teams are employed are business managerial staff, hospital emergency rooms, military command and control, congressional committees, and academic research (Hollenbeck, Ilgen, Sego, Hedlund, Major, & Phillips, 1995; Hollenbeck, Ilgen, LePine, Colquitt, & Hedlund, 1998; LePine, Hollenbeck, Ilgen & Hedlund, 1997).

In contrast, a heterarchy can be defined as a structure in which "members can freely exchange information with each other to jointly achieve the common goal" (Schraagen et al., 2010); essentially, the absence of a communications hierarchy (Walker et al., 2008). Heterarchical teams have been referred to as "flat" structure teams, "consensus decision-making groups" (Humphrey et al, 2002), or "network" teams in which decision rights are peer to peer (Schraagen et al., 2010). A heterarchy has been noted to be an emergent, self-organizing form that resembles "a network or a fishnet," in that it has lateral or distributed authority, and involve relationships of interdependence (Levchuk et al., 2004).

Environments in which organizations operate are becoming more unpredictable, and organizations are not always able to predict the extent of the demands they will encounter (Schraagen et al., 2010). In response, many organizations have adopted a form of a heterarchical, decentralized, distributed organization structure (DeSanctis & Jackson, 1994; Drucker, 1988; Lipnack & Stamps, 2000; Martins, Gilson, & Maynard, 2004; Priest, Stagl, Klein, & Salas, 2006). Recent military terminology has named these structures "power to the edge" structures (Alberts & Hayes, 2003).

The current research indicates, anecdotally, that a heterarchical structure may be a condition that promotes equal power by dividing up the work, which allows each member to attain some level of domain expertise.

Galbraith (1973) proposed a "contingency theory," that there is no "one size fits all" structure for organizations, and that it is dependent on the situation. While some research has shown that some hierarchical teams perform and adapt well to increased time pressure (Adelman, Miller, & Henderson, 2003), other studies have shown that the heterarchical structure resulted in better team performance and more communication than the hierarchical structure (Bowers, Urban, & Morgan, 1992).

Schraagen et al. (2010) found that heterarchical teams were faster and more accurate than hierarchical teams, attributing their accuracy to a better exchange and sharing of knowledge in the heterarchical structure.

1.2.5 Self-Organizing Teams

Many current heterarchical, network structured teams are characterized by their ad hoc, self-organizing nature. Many organizations are increasingly relying upon teams that are assembled swiftly, and able to address urgent and novel issues (Kozlowski & Bell, 2003). In many cases, these "temporary" teams are also self-organized teams. Highsmith (2004) states that self-organizing teams are composed of "individuals [that] manage their own workload, shift work among themselves based on need and best fit, and participate in team decision making."

Examples of such teams are large, multinational military coalitions that are assembled of people who may never have worked together before, nor are likely to see each other again after the mission is accomplished. Other examples are crisis management and emergency rescue teams that are assembled from police, fire brigades, and paramedics to deal with unexpected situations that require coordinated effort for a limited time span (Schraagen et al., 2010; Xu & Zhao, 2011). Or the example can be as simple as a classroom scenario in which teachers have students self-organize in teams to solve a problem or work on a project. In the intelligence community, self-organized teams occur frequently, as there is a need for analysts to work together, on many occasions from remote locations and disparate organizations, and share information to solve complex intelligence problems (Arney, Cohen, & Medairy, 2004).

Many self-organized teams can be temporary teams, and the time span for such teams can be extremely limited. In the case of an air traffic control scenario, the controller, pilot, and ground crew form a temporary team in order to successfully accomplish the task of operations within the airspace of the airport. This time period can be anywhere from a few minutes, from the time a local controller receives an aircraft until the time they depart, or up to twenty minutes for larger airspaces. These teams are indeed temporary in that even if the pilot goes through the same airspace again at a later date, it will be with a different controller, and so a different, temporary team.

Although the time duration of many self-organizing teams can be very short, members of these teams can come to be identified and committed to their teams (Chidambaram, Bostrom, & Wynne, 1990), and become a cohesive unit that can engage in collective action (Van De Ven, Van De Vliert, & Oosterhof, 2003).

Some have attempted to "organize" self-organizing teams by identifying six informal roles that were adopted in their study of self-organization: mentor, coordinator, translator, champion, promoter, and terminator (Hoda, Noble, & Marshall, 2010). However, it is recognized that roles can vary from situation to situation. And when confronted with these various scenarios, the power of self-organizing teams resides in people who will adapt and do what is required of the situation, thereby organizing themselves as a team to accomplish their common goal. This characteristic of self-organizing teams becomes a point of emphasis for the explorations of the current research. Team members are not recruited based on any prior experience or personalities, and exist together as a team for a relatively short period of time. By allowing the teams to self-organize, the research offers a naturalistic approach to teamwork and the effect of resources.

1.2.6 Team Size

In addition to team structure, team size is another important factor in teamwork and team performance. There is a "double edged" effect of team size. It may seem that adding more people as resources to solve a problem can relieve the load upon each individual, as there are more people to assist in the task. However, a larger team could both enhance and hinder decision making effectiveness and performance (Amason & Sapienza, 1997).

Larger groups can be more heterogeneous, which could be more conducive to creative problem solving. However, smaller teams can lead to increased team cohesiveness, improved communication, and coordination (Horwitz, 2005). In some studies, team size has been found to be positively related to cognitive and affective conflict (Amason & Sapienza, 1997). Larger size introduces diversity, which can benefit the team, but diversity can also introduce conflicts within the team (Shaw, 1981; Smith, Smith, Olian, Sims, O'Bannon, & Scully, 1994). As group size increases, the difficulties of agreeing on objectives, ensuring appropriate participation in decision making, achieving consensus on what constitutes high quality, and eliciting unanimous support for innovation, all increase. In larger groups, group members would be expected to be less engaged in the group, and can be less likely to perform with any concern for excellence (Salas, Rozell, Driskell, & Mullen, 1999).

Bradner and Mark (2002) found that members of smaller teams (4, 6, 7, 9) were more active participants, more committed, and more aware of team goals. But their results also found that in their study, larger teams (14, 15, 17, 18) paid more attention to coordination processes, in that they were better at maintaining formal procedures.

Some researchers have suggested that teams of three to five members are ideal because participation and coordination have greater chance of effectiveness than teams with more than five members (Bray, Kerr, & Atkin, 1978; Fern 1982; Hare, 1981; Shaw, 1981). But it is recognized that optimal size is highly task dependent.

There does exist an optimal size principle, which was demonstrated by Campion, Papper, and Medsker (1996). Their study showed that a group should be large enough to have enough resources to accomplish the required work, but not so large to induce difficulties in coordination. For example, in a study of operators using audio systems for distributed communication, Fincannon, Evans, Phillips, Jentsch, and Keebler (2009) found that teams of three, rather than two, were optimal for performance.

In another example, a dynamic decision task that mimicks an emergency response scenario, Guastello (2010) studied team sizes of 4, 6, 9, and 12. His study found that the optimal group size for the dynamic decision task was 9 player teams. His conclusion was that emergency response teams should be large enough to contribute a critical mass of ideas, but no so large that there is a compromise in coordination tasks.

Aubé, Rousseau, and Tremblay (2011) showed the negative relationship between team size and quality of group experience. Their study involved 97 work teams from a public safety organization, and show that overly large teams can produce counterproductive behavior and adversely influence their effectiveness.

In a study of the effect of team size on team coordination under time pressure conditions, Majalian, Kleinman, and Serfaty (1992) found that the average performance of triads was superior to that of individuals, who did slightly better than dyads. They attribute the success of triads to clear division of responsibilities, while dyads had greater uncertainty of coordination due to shared resource responsibility.

Researchers have suggested that teams are most effective when they have sufficient, but not greater than sufficient, numbers of members to perform the task (Guzzo, 1988; Guzzo & Shea, 1992; Hackman, 1990). Horwitz (2005) found that once team size goes beyond the optimum size, there seem to be increasing process losses and decreasing team integration that surpass whatever gains that can be had from a large team size.

Ultimately, it is recognized that an optimal team size may exist, but that it varies according to many factors inherent in teams and the natures of tasks that teams needs to accomplish (Horwitz, 2005; Guastello, 2010).

1.3 Performance Metrics

There are many metrics that have been developed to assess the described components of team performance and teamwork. Some examples are workload, accuracy, communication, error rate, reaction time, situational awareness (SA), task time, trust, usability, and many others (Billman & Steinberg, 2007; Entin & Entin, 2001). While it is acknowledged that there are many existing instruments to measure subjective assessments, an exhaustive survey of performance metrics will not be presented here. However, two subjective measures of SA and workload utilized in the research are discussed below.

1.3.1 Situational Awareness Measure: SART

Situational Awareness Rating Technique (SART) is a subjective rating scale that measures the participant's rating of his or her degree of situational awareness. This measure involves three main concepts: Demand of Attentional Resources, Supply of Attentional Resources, and Understanding. Each concept contains its own sub-concepts. The sub-concepts for Demand are Instability of Situation, Variability of Situation, and Complexity of Situation. The sub-concepts for Supply are Arousal, Spare Mental Capacity, Concentration, and Division of Attention. The sub-concepts for Understanding are Information Quantity, Information Quality, and Familiarity (Selcon, Taylor, & Koritsas, 1991; Taylor, 1990; Endsley, Selcon, Hardiman, & Croft, 1998).

For the SART ratings, each participant rates each sub-concept on a scale of 7 point rating scale (1=Low, 7=High), and the scores are then combined to form an overall SA score with the following formula:

$$\text{SA} = \sum \text{Understanding} - \left(\sum \text{Demand} - \sum \text{Supply}\right)$$

The full SART questionnaire can be found in Appendix K.

1.3.2 Workload Measure: NASA-TLX

Prior research has investigated the impact of workload on a variety of parameters related to team performance. Bolstad and Endsley (2000) investigated levels of workload and their impact on the formation of team situational awareness, and found that as workload increased, team interaction and direct communication decreased, resulting in an overall decrease in shared situational awareness.

The goal, however, is not necessarily to lower workload, but to understand the spectrum of workload and its impact on performance and SA. In fact, some studies have shown that small increases in workload resulted in increases in alertness and concentration (Verma, Kozon, Cheng, & Ballinger, 2007).

One of the most common measures of workload is the NASA Task Load Index (TLX), a multi-dimensional scale designed to obtain subjective workload estimates from operators. NASA-TLX was originally developed for use in aviation (Hart & Staveland, 1988), but has since been applied to a variety of fields and domains (Hart, 2006). It contains six subscales: Mental, Physical Demand, Temporal Demand, Frustration, Effort, and Performance. The descriptions of these subscales are shown in Table 1.2, and the NASA-TLX questionnaire that was used for this research is shown in Appendix K.

Table 1.2 *NASA-TLX definitions (Hart & Staveland, 1988)*

RATING SCALE DEFINITIONS		
Title	Endpoints	Descriptions
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

1.4 Workload

Evaluating workload is an important aspect of a system evaluation. Subjective workload measurement tools have been developed to quantify the effort a user exerts during the performance of a task. And although it has extensively been researched with individuals, the concept of workload has not been exposed to thorough investigation with respect to team performance.

Workload has been described as "the cost of accomplishing mission requirements for the human operator" (Hart, 2006), and mental workload as the "difference between cognitive demands of a particular job or task and the operator's attention resources" (Rubio, Diaz, Martin, & Puente, 2004).

Regarding mental workload, there has been little consensus on precisely what it is and how to measure it. Sheridan (1979) regards information processing and "emotional workload" to be the primary components of mental workload. Regarding workload measurement, it is his conviction that the most direct measure of mental workload is subject judgment, as compared to physiological or secondary task measurements. Sheridan's position is that mental workload is inherently subjective.

1.4.1 Workload and Performance

As mentioned, the reason workload is an important concept for team research is in regard to its relationship to performance. The relationship between workload and performance has, in some situations, been proposed to mimic the Yerkes-Dodson inverted-U relationship (Hancock & Ganey, 2003). This relationship indicates that there is an optimal level of workload that results in optimal performance, while both low and high levels of workload results in degraded performance (Cummings & Nehme, 2010). This is illustrated by the workload-performance curve inspired by the Yerkes-Dodson inverted-U relationship shown in Figure 1.1.

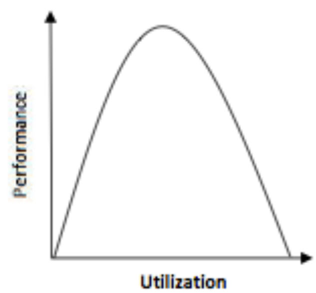


Figure 1.1 *Workload-Performance curve (Cummings & Nehme, 2010)*

Cummings and Nehme (2010) investigated this workload-performance relationship within a network centric warfare (NCW) concept in which operators will have access and will be required to take on more tasks and process more information than current operators. In this study, workload was measured as utilization, or "percent busy time," and was hypothesized to affect performance by mimicking the Yerkes-Dodson relationship, that performance can be degraded at both the high and low ends of the notional utilization curve. Therefore, identifying the areas of optimal performance as a function of workload, as well as identifying the thresholds for degraded performance were keys to designing efficient systems. The Cummings and Nehme (2010) acknowledge that true workload-performance curves are likely not parabolic in the symmetrical sense, and also point out the temporal and dynamic nature of utilizations as observations for future work.

It has been documented that the original Yerkes-Dodson experiment investigated the relationship between discrimination and strength of stimulus (Yerkes & Dodson, 1908). It is this original work that gave rise to the inverted-U relationship. Detailed investigation of this work reveals some interesting findings. Yerkes and Dodson (1908) performed experiments involving mice, not humans, and it was not focused on performance, but on differentiated rates of learning of mice. They performed three experiments aimed at investigating learning under different conditions of black/white discriminability. Learning was determined by how many days a mouse required to complete a successful discrimination 10 times in one day. Illumination level was varied to make discrimination of black and white at three levels. Shock was delivered to the mouse once it entered the black area, and the time it took to learn to go to the white area was measured.

Hancock and Ganey (2003) point out that Yerkes and Dodson themselves did not connect their results with stress and performance. He quotes Teigen (1994), who says, "to the animal experimenter of 1908, speed of habit-formation is speed of habit-formation and nothing else. The tasks vary in 'difficultness of discrimination,' and strength of shock is simply 'strength of stimulus' with no attempt to speculate about its aversiveness, or its emotional or motivational significance." However, despite the criticism, Hancock and Ganey (2003) notes that the description of the inverted-U has "a strong intuitive appeal," and although the original experiment does not appear to have exact relevance to workload and performance, the general spirit of the parabolic relationship has been invoked for subsequent investigations (Hebb, 1955; Cummings & Nehme, 2010), aimed at identifying the areas of optimal performance as a function of workload, as well as identifying the thresholds for degraded performance, towards the goal of designing efficient systems.

1.4.2 Workload and Situational Awareness

One important component to team decision making and performance is attaining optimal situational awareness (SA) (McCue, 2005; Adams, 2007). There are a variety of factors that can influence SA, but one parameter that has a significant effect is workload. Therefore, the goal then becomes to optimize both SA and workload, and understand the relationship between the two (Endsley, 1993). It is evident that in a high workload situation, a person's SA would be degraded (Scerri, Owens, Sycara, & Lewis, 2010). Research has also shown that decreased SA can result from both high and low workload situations (Hendy, 1995), similar to the workload-performance relationship mentioned above (Cummings & Nehme, 2010).

This relationship has particular relevance to the intelligence community, where teams operate together to make decisions to uncover clandestine networks. Obtaining optimal team workload and SA by establishing appropriate roles and number of analysts could provide valuable guidance in organizing these teams.

Hendy (1995) explored this relationship between workload and situation awareness, and how performance is dependent on these two concepts through the framework of the Information Processing (IP) Model and William T. Power's Perceptual Control Theory (PCT) (Powers, 1973).

The IP Model shows that workload and performance are driven by the ratio of "time taken to process the information necessary to make a decision" to "time available before the decision has to be actioned" (Hendy, 1995). The ratio essentially provides a measure of the time pressure, and the IP model proposes that performance, errors, and subjective workload are all determined by time pressure. The IP Model predicts adaptive operators, when confronted with excessive time load, will either reduce the amount of information to be processed, or increase the time before a decision needs to be put into action. Degraded performance would occur if either of these actions is not taken. One factor in reducing amount of information to be processed is knowledge, which is described by Hendy (1995) as "resolved uncertainty." This leads to discussions about situation awareness (SA) and mental models. The mental model contains "the operator's goal state and the reference against which actions are selected and initiated." And so it is the mental model that plays a key role in shaping perception and action in goal-directed human activity.

One of the key claims of Perceptual Control Theory (PCT) is that perception is what is controlled, not behavior (Powers, 1973). An illustration of PCT is shown in Figure 1.2.

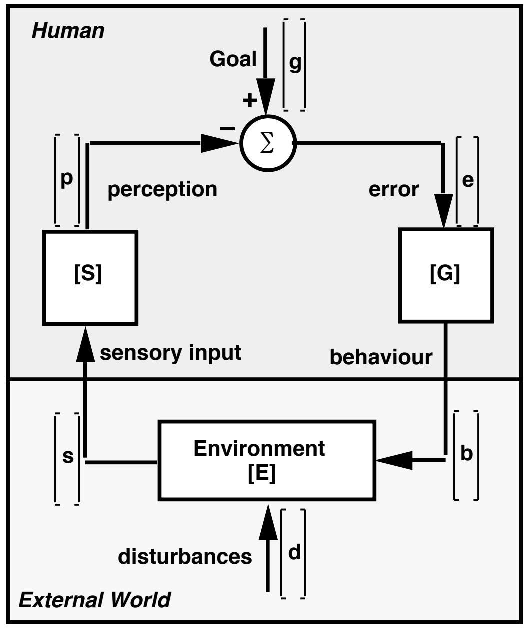


Figure 1.2 *Diagram of Perceptual Control Theory (Hendy, 1995)*

As diagrammed in Figure 1.2, Forssell (2008) provides a description of the fundamentals of PCT, which is an extension of basic control theory, and mimics a simple closed-loop control system. As shown in the figure, there exists a reference signal, or goal state, that specifies the state to which the perception must be brought. The reference can be thought of as a goal, the state of something to be experienced, and is set from inside the person as a whole, not by sensory inputs. The reference signal enters the comparison function, which then outputs an error signal (the difference between the reference signal and what the person is experiencing right now). Humans likely have a stronger sense of the error signal than the reference signal. The error signal goes to the output function, which processes the error signal. The output quantity is something physical, an action, or most commonly referred to as behavior. The effect of the action of the controlled variable becomes the feedback function. And so behavior becomes the control of perception, taking into account disturbances to the environment (Forssell, 2008).

Team performance, as the current research indicates, can be viewed from the perspective of PCT. Each team, and each team member has a goal state, performance. As each team member compares his/her performance in relation to the team's performance, an error signal is generated, which provides feedback to that team member. Using the feedback, each team member correspondingly adapts his/her behavior to control for their perception of overall team performance. Each team member's perception of team performance with respect to their own behavior then becomes an important component to a team's overall performance.

Examination of the diagram shows how SA and workload can be related. SA is gained over time through interaction with the environment. But in a changing environment, the development of SA is a task that demands attentional resources. So the two concepts are related in that the development of these aspects of the mental model depends on the availability of processing resources for active control of these processes.

Hendy (1995) describes a simulated Air Traffic Control experiment designed to investigate the relationship between workload and situational awareness, measured subjectively by NASA-TLX and SART. Principal component analysis (PCA) was used to determine factor loadings, how much of the variance can be attributed to an underlying feature, with the first principal component accounting for the greatest amount of variance. Using PCA with NASA-TLX and SART, Hendy (1995) examined the underlying relationships between the different components of the two measures to see what main factors may emerge (Pett, Lackey, & Sullivan, 2003). Two main factors emerged: a demand or workload-related factor, and a factor related to acquired knowledge (SA factor) (Hendy, 1995). These factors demonstrate that workload and SA are interrelated, and that both factors independently, as well as jointly, need to be considered in understanding their impact on team performance. The current research does not investigate this joint effect, but recommends the study for future work.

Verma, Kozon, Cheng, and Ballinger (2007) investigated the impact of a new automation tool on procedures, roles, and responsibilities for air traffic control tower controllers. With the new automation tool, their hypotheses were that workload would become more equally distributed, situation awareness would remain the same, and voice communication loads would be more equally distributed. The participants were four retired controllers. NASA-TLX was used to measure workload and SART was used to measure situation awareness.

The results show that workload did spread out more evenly due to re-distribution of jurisdiction, as hypothesized. However, the hypothesis about situation awareness was not upheld, as it was shown to improve. This was explained partly by the interaction of workload and situational awareness. The spare mental capacity could bring about better stability and better awareness. Lastly, the number of communications decreased as overall communication became more efficient. These results demonstrate the dramatic impact that a change of workload has on the time a controller spends talking, indicating more efficient communication.

The study also hinted at behavior akin to the Yerkes-Dodson relationship, that small increases in workload resulted in increases in alertness and concentration, showing that the small increases in workload could have had the beneficial effect of preventing tedium and vigilance decrement (Verma, et. al, 2007).

Bolstad and Endsley (2000) examined the effects of two different types of shared displays and varying workload levels on the formation of team situational awareness (SA). The study built upon previous research that showed the benefit of shared displays for building shared mental models and improving team performance (Bolstad & Endsley, 1999). This study explored an abstracted shared display, which showed only the critical information from the other team members' display, instead of a complete replicate of the other members' display. The hypothesis was that the abstracted shared display would help build team SA without imposing the extra workload that comes from full shared displays.

The experiment comprised of 36 participants, tested in pairs. Two independent variables were "Workload Level" (within team manipulation) and "Shared Display Type" (between team manipulation). Three workload levels and three display conditions were tested for the "Theatre Defense program" task, a command and control type task. The results showed that the difference in display conditions on performance became evident under the high workload condition, and that as workload increased, team interaction and direct communication decreased, resulting in an overall decrease in shared situational awareness. It showed that the abstracted shared displays provided the most benefit to team coordination under high workload conditions when direct communication became difficult.

1.4.3 Team Workload

Today, few systems are designed and produced without evaluations of operator workload, underscoring the importance of this factor. However, Bowers, Braun, and Morgan, Jr. (1997) note the lack of research related to workload of teams.

Conceptually, the idea of "team workload" is derived from adapting individual workload concepts to the team environment. A central concept to understanding workload is the realization that task performance is a function of the cognitive resources dedicated to accomplishing a task. This function could be described as a relationship between the cognitive availability and resources required for the situation. When total quantity and demand remain constant, performance becomes a function of resource efficiency, which can be improved through practice and training.

When applied to a team, workload can be characterized as “the relationship between the finite performance capacities of a team and the demands placed on the team by its performance environment” (Bowers et al., 1997). According to this definition, team performance is optimized when team resources are in balance with situation demands. In a team setting, members are required to be engaged in two broad categories of activities: taskwork and teamwork. Thus, this becomes the differing point between individual and team workload, that the team experiences overall demands that go beyond the sum of the workload of individual members.

Based on these assertions, clearly defining the components of taskwork and teamwork become critical in determining the components related to team workload. Salas, Cooke, and Rosen (2008) define taskwork as “the components of a team member’s performance that do not require interdependent interaction with other team members,” and teamwork as “the interdependent components of performance required to effectively coordinate the performance of multiple individuals.”

However, the challenge lies in the fact that teamwork behaviors vary with task demands, and so may not be generalizable. It can be hypothesized that increases in demand associated with either taskwork or teamwork will result in increased workload and diminished performance, and furthermore, this workload can be diminished as a function of training, experience, and practice. Some factors that can influence team workload are coordination, communication, team experience, and team training.

In terms of the assessment or measurement of team workload, Bowers et al. (1997) suggest that subjective ratings are more conducive to the team environment than physiological indices. However, the issues that need to be addressed in assessing team workload are in determining how to combine data from individuals of a team, whether to require individuals to rate their perception of overall team workload, and determining what components of workload should be measured and how these components should be combined.

Prior investigations have used a modified NASA-TLX scale, but the results indicate that it lacks the sensitivity to capture team workload. The same observation is made for secondary task data, showing the apparent ineffectiveness of adapting existing measures of individual workload for use with teams.

Hildebrand, Pharmer, and Weaver (2003) used a modified version of the NASA-TLX in an attempt to measure team workload, called the "Team Workload Scale." Their study involved the Navy's attempt to optimize manpower aboard its ships through a prototype automation tool. Team performance, situational awareness, and workload were evaluated from a performance and usability evaluation on the automated task management concept for the Air Defense Warfare (ADW) project. Three hypotheses were proposed for teams using the automated interface: teams would perform more timely and accurately, teams would demonstrate improved situational awareness, and teams would exhibit more manageable and evenly distributed workload. Six teams (five members each) were tested on the ADW scenario. There were two segments to the simulation, of differing difficulty. Task outcome was measured through a performance index that was developed through subject matter expert (SME) input. SA was assessed through observational inferences as well as through a questionnaire. Participants filled out questionnaires in between the two segments, as well as after the simulation. Each participant rated their workload across ten dimensions: six from the original NASA-TLX, and an additional four dimensions to measure workload associated with interacting in a team environment. These four dimensions were demand for communications, monitoring, control, and coordination. In addition, a second workload measure was collected every ten minutes by SME evaluators. The results of the experiment did not completely support the hypotheses. There was no statistically significant difference in the SA and workload results. One interesting result was that in low difficulty situations, teams perceived the tool to hinder their performance, but did perceive it as an aid in high difficulty situations. There was not a significant difference in performance while using the tool versus without it. But this conclusion was contrary to the subjective assessments of the tool, in that overall, participants liked the new display and they felt that it helped them do their jobs. One of the benefits they felt was that the new display lessened the required communications.

Entin and Entin (2001) describe an additional method of capturing team workload, an extension of the NASA-TLX. The first part of their questionnaire captures the individual assessment of their own workload. The second part estimates the overall workload experienced by each of the other team members. The third part is a subjective assessment by each team member for the team as a whole.

Although these studies have all extended the understanding of team workload, the concept of team workload still requires much advancement, and a reliable measurement of team workload is as yet undeveloped and recommended for future work. However, the distinction of taskwork and teamwork (Salas et al., 2008), as discrete components of a team's process, are leveraged in the understanding of the team's progression in the ISR "Red/Blue" exercise, and will be further discussed in Chapter 2, section 2.3.3.

1.5 ISR "Red/Blue" Exercise

The apparatus for this research is the Intelligence, Surveillance, and Reconnaissance (ISR) "Red/Blue" exercise, a network discovery simulation game developed by the Intelligence and Decision Technologies Group at MIT Lincoln Laboratory (Won et al., 2011). The exercise was originally designed as a process for discovering new intelligence exploitation concepts, system designs, and algorithms for a broad range of ISR missions through competitive game play. Counter-insurgency operations require analysts to gain an understanding of a very complex scene; insurgents often conduct activities in the open, relying on the vast amount of urban "clutter" to disappear into background of the general population. The understanding requires that humans provide much of the information processing capability and understanding to maintain awareness and interpret intent of various activities (e.g., surveilling a possible target) as opposed to the activity itself (e.g., stopping at a market).

“Red/Blue” experimentation, in general, is a human-in-the-loop process to hone tools, workflows, and organizational concepts through a highly instrumented and competitive game, which is focused around a specific challenge area. Within a problem area, a relevant scenario is developed, along with a concept of operations (CONOPS) for the resources teams will have available to them (e.g., specific sensor types). As teams work through the scenario, measurements are taken on the use of the exploitation of tools provided to them, as well as observational data about behavior and organization. This information is then analyzed and fed into the tool development process. This “Red/Blue” process allows the involvement of the user in the development of tools to improve performance, as well as observe impacts of human factors issues in team performance.

In the ISR "Red/Blue" exercise, a "Blue" team works together to discover a complex hidden insurgent network ("Red" network) within a simulated urban environment. The team uses wide-area persistent surveillance data and other sources of information to trace relationships between individuals, events, and sites to uncover this hidden insurgent network. Although the data is simulated, it emulates a "real-life" command and control scenario in which a team coordinates and collaborates in order to solve a realistic problem.

The primary tool used to interact with the data is a tool called "BlueStreak", developed at MIT Lincoln Laboratory. It is a geospatial and temporal representation of all of the information available in the exercise. A screenshot of the map view of "BlueStreak" is shown in Figure 1.3.

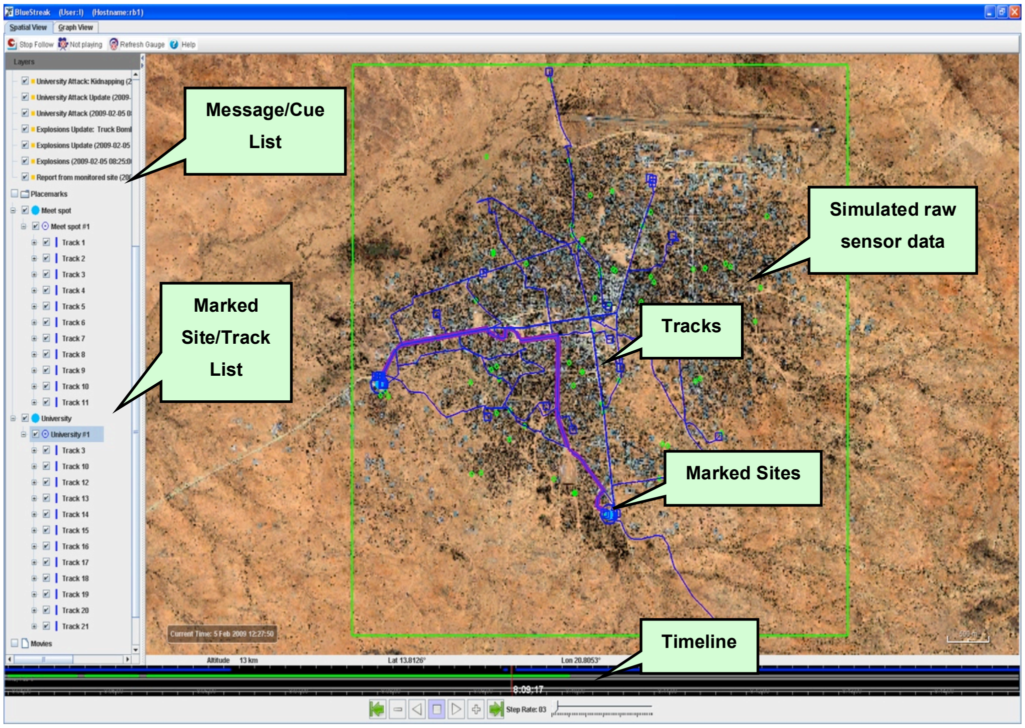


Figure 1.3 *ISR "Red/Blue" exercise software tool: "BlueStreak"*

As shown, "BlueStreak" provides a timeline and a spatial map view of the scene with an overlay of the sensor data. Users have the ability to view the data forensically, and simulated sensor data shows the movement of vehicles (tracks), which can be combined with information about various events (message/cues) to identify potential hostile sites (Won et al., 2011). Chapter 2 will describe the tool in more detail.

1.6 Information Theory

Information theory traces its beginning to Claude Shannon's 1949 paper, "A Mathematical Theory of Communication" in the Bell Systems Technical Journal, in which he demonstrated how information could be quantified. In the article, information is defined as "reduction of uncertainty" (Shannon & Weaver, 1949). Essentially, prior to an event, a person is more uncertain about the state of the world than after the event. When the event occurs, information has been conveyed, unless the event is completely expected (Wickens & Hollands, 2000).

Shannon proposed a way to measure and compare information, and proposed the idea to measure the quantity of information by considering its predictability, or probability. The notion is that messages could be compared according to how probable or improbable they are. When a message is entirely certain (probability = 1), the quantity of information conveyed is zero. On the flip side, if it's highly improbably (probability ~ 0), the quantity of information is very high.

Information theory comprises many other components and is covered in much greater detail in Shannon's article, as well as in the many applications of his theory in a variety of fields. However, for the purposes of this research, the primary importance of the theory is its premise that information is the "reduction of uncertainty." All transactions involving information, including information production, synthesis, sharing, and processing, will all be viewed from this perspective of humans as active information seekers, pursuing the reduction of their uncertainty in the given situation.

1.7 Signal Detection Theory

"Signal detection theory is applicable in any situation in which there are two discrete states of the world (signal and noise) that cannot easily be discriminated. Signals must be detected by the human operator, and in the process two response categories are produced: Yes (I detect a signal) and no (I do not)" (Wickens & Hollands, 2000).

At a simple level, the "Red/Blue" exercise is a game that evaluates how well teams of people can make a "yes" or "no" decision, given various stimuli. As such, the framework for understanding teamwork and team decision making in the game is based on Signal Detection Theory (SDT). In SDT, given that there are two states of the world and two potential responses, there are four possible events: hit, false alarm, miss, correct rejection, as shown in Figure 1.4.

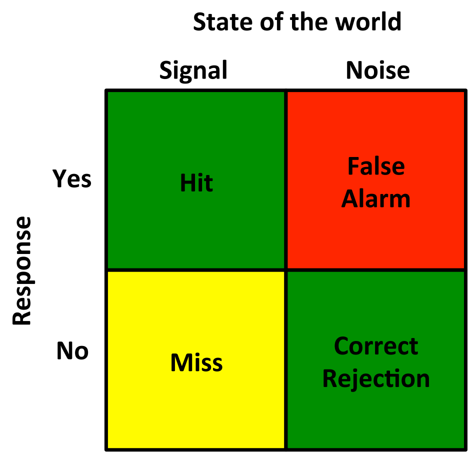


Figure 1.4 *Four outcomes of Signal Detection Theory*

Signal Detection Theory (SDT) first emerged from problems related to detecting weak signal in noise (Tanner & Swets, 1954), with the original framework involving the problem of a radar operator deciding whether what he sees on the radar screen indicates the presence of a plane (signal) or the presence of noise (Green & Swets, 1966; Sorkin, 1999). It has since been extended to describe a variety of sensory and perceptual behavior (Macmillian & Creelman, 1991; Wickens, 2002). The essence of SDT is the measurement of a subject's ability to detect or discriminate sensory stimuli, and provides the language and concepts (see Figure 1.5) for analyzing decision making (Heeger, 1997).

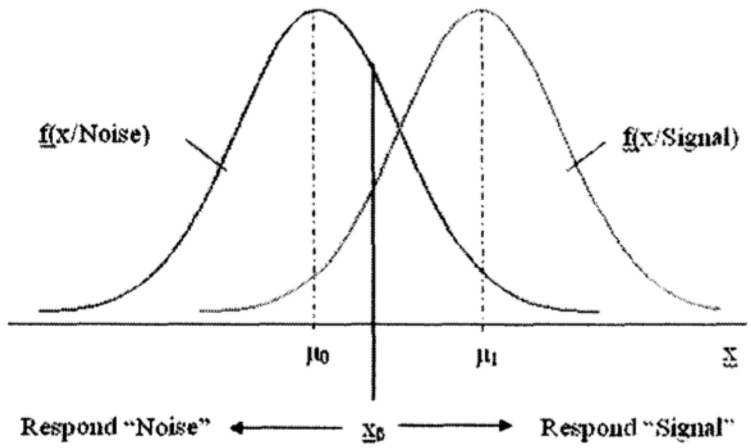


Figure 1.5 *Signal and noise distributions (Green & Swets, 1966)*

As Figure 1.5 shows, given a signal detection task, SDT typically assumes that there are two distributions. The curve on the left shows the distribution when there is no signal present (noise only), and the curve on the right is the distribution when there is a signal present (signal-plus-noise). It is assumed that these distributions are equal variance, normal distributions. SDT assumes that the signal is received in the context of noise. Therefore, the signal-plus-noise distribution is the noise distribution displaced by the amount the signal differs from the mean of the noise. When the subject receives a stimulus, the task is to determine if the strength of the signal is sufficiently different from the noise distribution to conclude that the signal was present. Performance is then described as the subject's ability to discriminate signal from noise. As shown in Figure 1.5, this ability is described in SDT as "sensitivity", or *d'*, which is the normalized distance between the means of the two distributions (Sorkin, 1999), where μ_1 represents the mean of the signal-plus-noise distribution, μ_0 is the mean of the noise distribution, and σ is the variance:

$$d' = \frac{\mu_1 - \mu_0}{\sigma}$$

As *d'* increases, it means that the distance between the means are greater, resulting in a smaller overlapping region, increasing the ability of the subject to discriminate the signal from the noise. The value of *d'* is typically greater than 0, where $d'=0$ represent chance performance (inability to distinguish signal from noise), and higher the *d'*, the better the sensitivity becomes. In other words, the higher the sensitivity (*d'*), the subject will have a higher ratio of hits to false alarms.

The probability that the subject responds "signal" given an input from the signal-plus-noise distribution is typically referred to the hit probability, *P(H)*, and the probability that the subject responds "signal" when the input is from the noise distribution is the probability of false alarms, *P(FA)*. Lastly, the vertical line in Figure 1.4 corresponds to the decision criterion of the subject, represented as x_c. A similar way to consider the criterion is a variable that is positively correlated with x_c:

beta (β), which reflects the bias of the subject in making a "signal" response. If the criterion were located at the intersection of the two distributions, it would indicate the decision criterion of an unbiased observer. The criterion essentially identifies the point at which the subject responds "yes" (to the right of the line) or "no" (to the left of the line). In the example of Figure 1.4, the subject has adopted a decision criterion to the left, which reflects a bias in favor of saying "signal". By doing so, this subject is willing to increase their probability of hits, *P(H)*, which is the area under the signal-plus-noise curve from the criterion point, while accepting a greater probability of getting false alarms, *P(FA)*, which is the area under the noise distribution curve from the criterion point. These probabilities are shown in the following equations.

$$P(H) = P(\text{"signal"} \mid S + N) = \int_{x_c}^{\infty} f(x \mid S + N)dx$$

$$P(FA) = P(\text{"signal"} \mid N) = \int_{x_c}^{\infty} f(x \mid N)dx$$

P ("signal" / S + N) is the probability of detecting a "signal" given the signal-plus-noise distribution, and *P ("signal" / N)* is the probability of detecting a "signal" given the noise distribution. When these probabilities are equal, $\beta = 1$, which indicates no bias.

$$\beta = \frac{P(\text{"signal"} \mid S + N)}{P(\text{"signal"} \mid N)}$$

When $\beta > 1$, the bias is towards saying "no" when there is a signal present, and when $\beta < 1$, the bias is towards saying "yes" when there is a signal present.

These basic principles of SDT will be used to quantify and measure team performance and teamwork by measuring a team's ability to accurately detect and act on signals while successfully filtering out noise.

1.8 Research Objectives

The proposed research investigates the aspects of teamwork and team performance and the impact of resource allocation in the context of a two-stage decision making process in an ISR simulation exercise. Signal Detection Theory (SDT) is the framework for modeling the teamwork and performance parameters, and the research process is from an ecological approach (Gibson, 1972, 1979).

Ecological psychology originated from James Gibson (1972, 1979). It aims to study human-environment relationships and human perception in rich, complex environments. It is contrasted to the more traditional experimental psychology, in that experimental psychology studies human behavior primarily in laboratory environments, and to a certain extent, with perceptual and cognitive psychology, which is less focused on human-environment interactions (Burns & Hajdukiewicz, 2004). Ecological psychology advocates that human behavior is often constrained by the environment that humans work in (Gibson, 1979). Because the ultimate goal is to design for humans working in complex environments, not laboratories, and we know that their environment presents them with certain constraints to work within, ecological psychology provides a useful approach to study teamwork in the environment provided for this research in the ISR "Red/Blue" exercise.

Therefore, this research proposes to investigate teamwork and team performance through the naturalistic behavior of teams in different team organizations that are promoted by resource allocation. Again, much teamwork research has either involved fabricated laboratory settings or complete field observations where manipulations are difficult. The present research aims to allow teams to behave naturally within team organization structures that are promoted only by the number of resources given them, namely, computers. The aim is not to specifically examine factors such as types of displays (Bolstad & Endsley, 2000), leadership style (Solansky, 2008; Somech, 2006), or training style (Salas, Dickinson, Converse, & Tannenbaum, 1992), to name a few examples of previous team research. While these types of research have their own merit, the effort of this research is to push the envelope of what naturally occurs in team organizations simply through resource allocation. To that extent, teams are allowed to self-organize. The intent is to allow roles to naturally emerge in a team, dictated by the available resources. As such, specific restrictions and rules are not placed on the teams, such as communication and interaction methods. The affordances provided by the available resources are what will drive the mechanisms for communication and collaboration that mark particular team concepts, and the corresponding results of teamwork and performance.

Furthermore, while it has been acknowledged that SDT provides the proper framework for understanding decision making in uncertain situations (Heeger, 1997; Wickens, 2002), much of the research has involved individuals, and application of SDT to understand team decision making has not been thoroughly explored. This research also proposes an innovative approach to teamwork, by decomposing team decision making to a two-stage process within the SDT framework. Team sensitivity (d'), which quantifies the team's ability to distinguish signal from noise, will be measured at both stages and used to describe teamwork. This two-stage process is further described in Chapter 2.

This research was conducted in a two-phase approach. The first phase of experiments, Phase 1, was conducted to explore the impact of team size and resources (e.g., number of people and number of computers) on performance. The results from the first phase gave rise to an emergent teamwork parameter, which distinguished higher performing teams from those who did not. These results and observations led to the second phase of experiments. In the second phase, Phase 2, the number of people per team was held constant, and the number of computers was varied. This manipulation was to study the impact of resource allocation on how teams would naturally organize themselves, how the organization would impact teamwork and performance. It is understood that teamwork is sufficient, but not a necessary condition for good performance. But it is also understood that high levels of teamwork can, on a more consistent basis, lead to high performance. Therefore, the research aims to identify a potentially necessary and sufficient condition of resource allocation that promotes teamwork.

2

General Method

2.1 Participants

The experiments were performed in two phases, Phase 1 and Phase 2. Participants for Phase 1 were recruited through target of opportunity activities at MIT Lincoln Laboratory, as well as the student subject pool at Tufts University. Participants for Phase 2 were recruited from the student subject pool at Tufts University. Sample sizes for Phase 2 were determined from power analyses of Phase 1 data. For the 4-person, 6-person, and 8-person teams for Phase 1, participants were recruited through targets of opportunity from workshops conducted at MIT Lincoln

Laboratory, and are considered part of the exploratory phase. Sample sizes for individual participants and 3-person teams for Phase 1 were determined from power analyses of pilot data. All participants were English speaking and comfortable operating a computer.

2.2 Apparatus

The experiments were conducted on Dell Precision M6500 laptop computers, running Windows XP (64-Bit), with 17-inch screens and a screen resolution of 1280 x 1024 pixels, 24-bit color. In addition, office supplies, a projector, a whiteboard, and a large printed map of the geographic area were provided. The setup at the Tufts University Human Factors and Applied Cognitive Systems Engineering Laboratory is shown in Figure 2.1.

Figure 2.1 *Experiment room setup*

As mentioned in Chapter 1, the principal apparatus for this research was the ISR “Red/Blue” exercise, a network discovery simulation game developed at MIT Lincoln Laboratory. In the game, the participants comprise the “blue” team, who work together to discover a hidden insurgent network (“red” network) within a simulated urban environment. The team uses wide-area persistent surveillance data and other sources of information to trace relationships between individuals, events, and sites to uncover this hidden network. Each team is allowed a set of tools, including laptop computers, projector, white board, a physical geographical map, and office supplies, which can all be see in Figure 2.1.

As mentioned in Chapter 1, the primary software tool used to interact with the data is called "BlueStreak", developed at MIT Lincoln Laboratory. Figure 2.2 shows a screenshot of the "Bluestreak" map view, which is a geospatial and temporal representation of the information available in the exercise. Simulated sensor data shows vehicle tracks (where vehicles came from and went to), which can be combined with information about various events to identify potential hostile "red" sites (Won et al., 2011). As vehicle tracks and potential sites are marked as sites or tracks of interest by the user, they appear graphically on the map and are also listed in the panel on the left side of the screen.

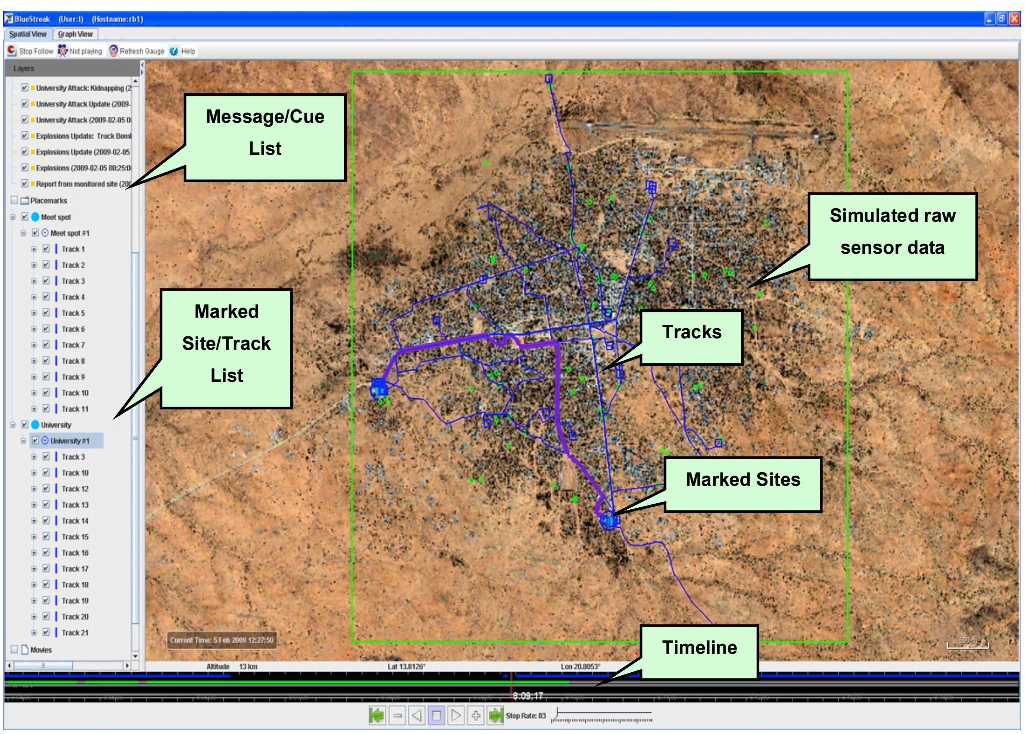


Figure 2.2 *"BlueStreak" map view*

A parallel "graph view" is also provided in the tool to assist the team in organizing their information in a graph structure (see Figure 2.3). The graph view is not required to be used, but it can allow the user to establish nodes and links for the sites of interest in a graph structure via the computer interface, it they so choose.

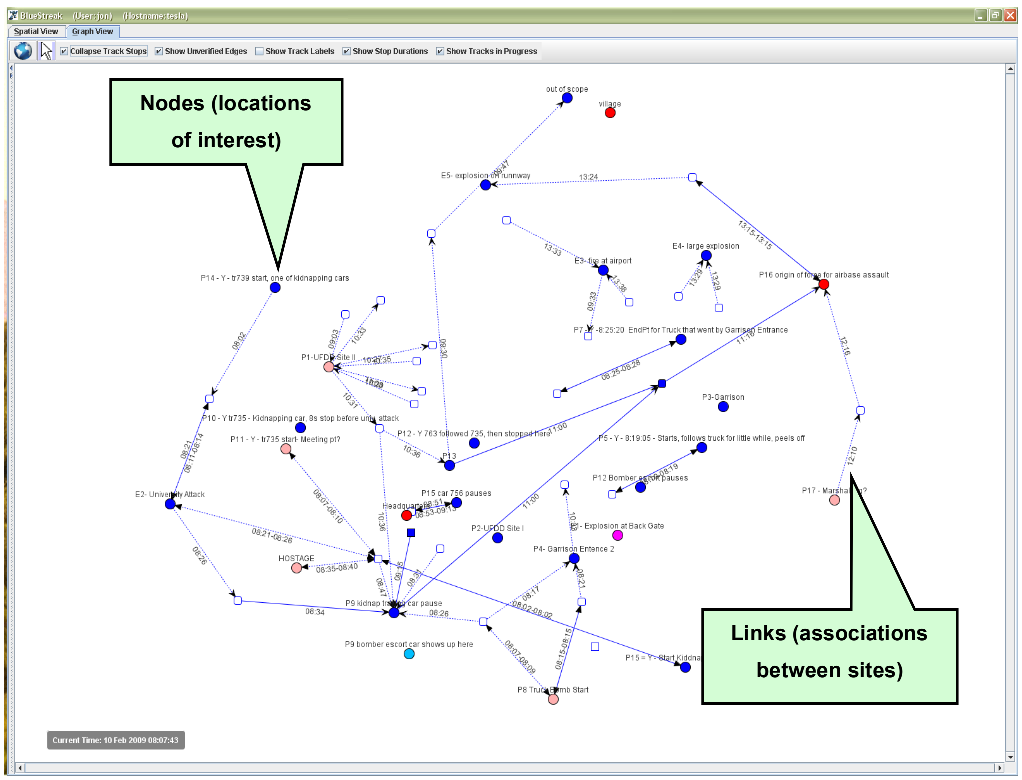


Figure 2.3 *"Bluestreak" graph view*

The "Bluestreak" tool is collaborative in nature, and all annotations and tracks generated on the map and manipulations of the graph are all shared among team members. Each laptop computer has access to the same data, but the data that is shown in the viewport on each computer is customizable by the user.

The full duration of the session is two hours. Following is an outline of the experiment.

1. Preliminary questionnaire (gender, age, computer experience, video game experience, intelligence analysis experience)
2. Introduction to "Red/Blue"
3. Training to Criteria
4. Mission Brief
5. Training Scenario
6. Team Planning Session
7. Game Play
8. Postgame questionnaire (subjective assessments)

Teams are given an introductory briefing for 5 minutes, describing the ISR "Red/Blue" exercise and the concept of network discovery. Following the introduction, teams are trained to criteria on the tool. Teams are trained to perform all the items on a checklist of criteria that have been identified as fundamental skills to operate the tool. The full training to criteria checklist with descriptions is provided in Appendix M. This training time takes place for approximately 15 minutes. Teams are then given a mission briefing for 5 minutes, in which they are given the game objectives and scoring scheme. Teams then play a training scenario, in which they have further practice with the tool, and are given an opportunity to strategize and distribute roles and responsibilities. The training scenario takes place for 25 minutes. After the training, teams then play the actual game for 1 hour. Following the game, all participants are given questionnaires that measure their subjective experiences of the game.

2.3 Procedure

The goal of the ISR "Red/Blue" game is to uncover a simulated clandestine insurgent network in the city of Abeche, in the country of Chad. The teams are told that there are troops on their way to Abeche to provide security and operations support, and that they, as the "blue" team of analysts, need to analyze the data and provide an estimate of all the "red" locations. All of the data analysis is performed forensically, meaning that all of the data, messages and cues are given from the beginning (e.g., no new data is introduced during the game). From their identified "red" locations, they are to make recommendations for the incoming troops to "surveil" or "assault." These are the only two action recommendations that are required of the team, and they will be scored based on these recommendations of actions. Lastly, of the sites that teams recommend, they are encouraged to identify the most likely red headquarters. If correctly identified, they will receive bonus points, but will not be deducted any points for mislabeling it.

During the scenario brief, teams are given an overview of these game objectives as a task force command (see Appendix I), as well as the scoring matrix (see Appendix J).

2.3.1 Clandestine Network Construction Workflow

There are a variety of strategies that teams can employ for playing the game and building their picture of the clandestine network. A typical workflow for the network construction task is described below in Figure 2.4.

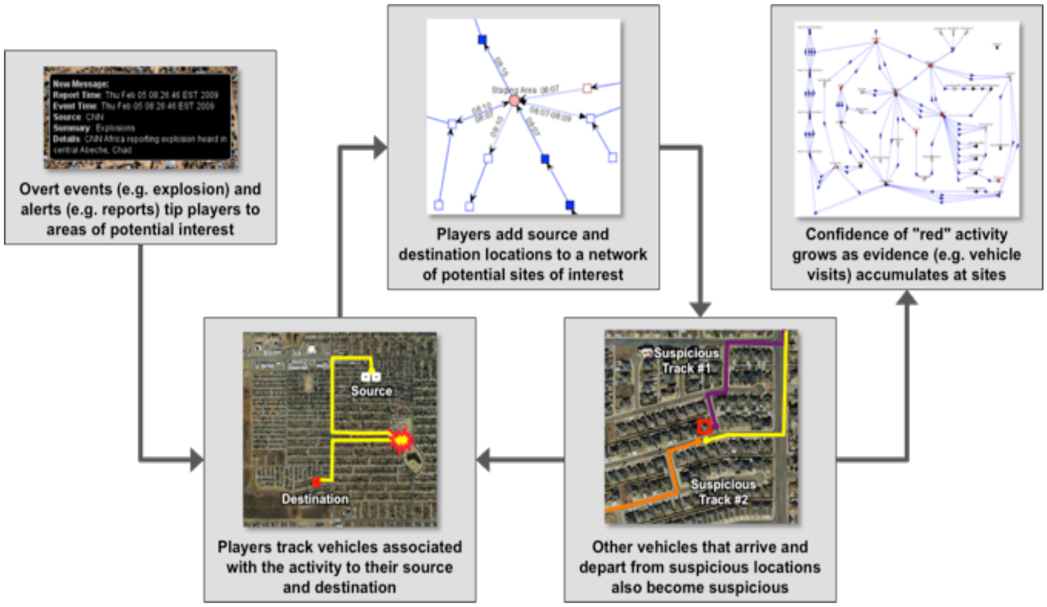


Figure 2.4 *Diagram for a typical network construction workflow*

Teams are provided messages regarding overt events or tips. These messages provide cues to various locations of interest. From a given location of interest, one can track vehicles that came to that location, or departed from that location, while paying attention to the times of the events or cues, as well as vehicle arrival or departure times.

As teams begin to track vehicles to and from various locations, they can employ one of two basic strategies, “breadth-first search” or “depth-first search,” which are graph search terminology from graph theory. Nodes and branches, from graph theory, as applied to the “Red/Blue” game are locations and tracks.

In a breadth-first search, starting from one location, teams would look at all the tracks from that one site and all the sites that all those tracks go to. Then for each of the nearest sites, teams explore all the tracks from that site, and continue that process. An example diagram of a breath-first search is shown in Figure 2.5, where the numbers on the nodes represent the order in which the nodes would be expanded.

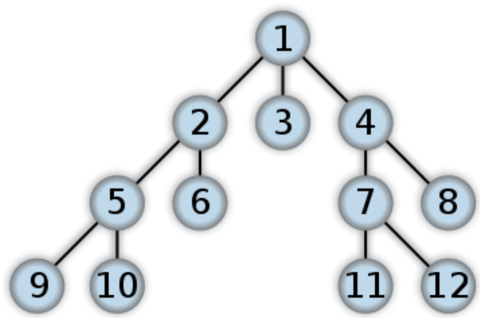


Figure 2.5 *Example of breadth-first search*

In a depth-first search, teams would start from one location and explore one series of tracks as far as possible before backtracking. An example diagram of a depth-first search is shown in Figure 2.6, where the numbers on the nodes represent the order in which the nodes would be expanded.

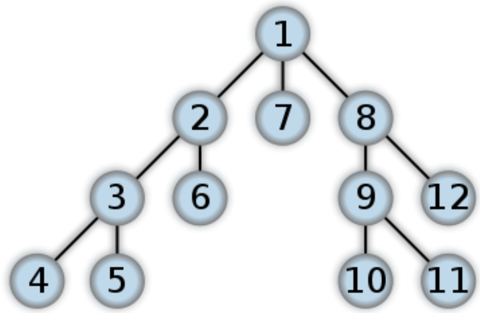


Figure 2.6 *Example of depth-first search*

Using either of the strategies, or a combination of the strategies, teams build out their network. As suspicion of certain vehicles grow, the sites that these vehicles visit become suspicious as well. Through the building out of the network, suspicion of certain sites as hostile "red" sites grow, ultimately leading to the recommendation of action at these red sites.

2.3.2 Detailed Task Description

This section provides a more detailed description of the workflow described above. The following describes the steps of a typical user who is operating one of the computers. There are no typical tasks for members who are not operating a computer. As mentioned previously in the chapter, teams are provided other resources, such as pen and paper, a whiteboard, a large physical map of the city, and an LCD projector with a switch that can project any one of the laptop screens. Team members who are not operating a computer could notionally use these other resources, if they so choose.

At the beginning of the game, users log on to the initial map-view, shown in Figure 2.7. The green dots represent the raw sensor data of the corresponding vehicles. The green boundary represents the field of view of the sensor, and as such, there is no data outside of the boundary area. Teams are informed that they will not see any data presented outside of the box. In addition, teams are also informed that the simulated sensor which is generating this data has enough resolution to detect vehicles, but not people, and so all of the data presented in the game will be vehicle data.

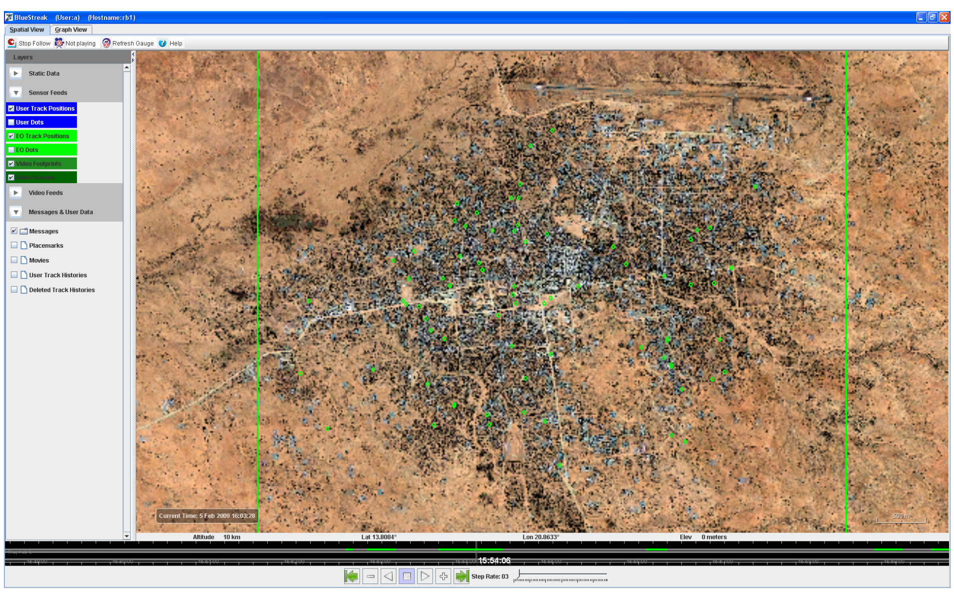


Figure 2.7 *Initial map view*

The typical first step is to view all of the messages that are present for the game, by double-clicking on the "Messages" label at the left panel, shown in Figure 2.8.

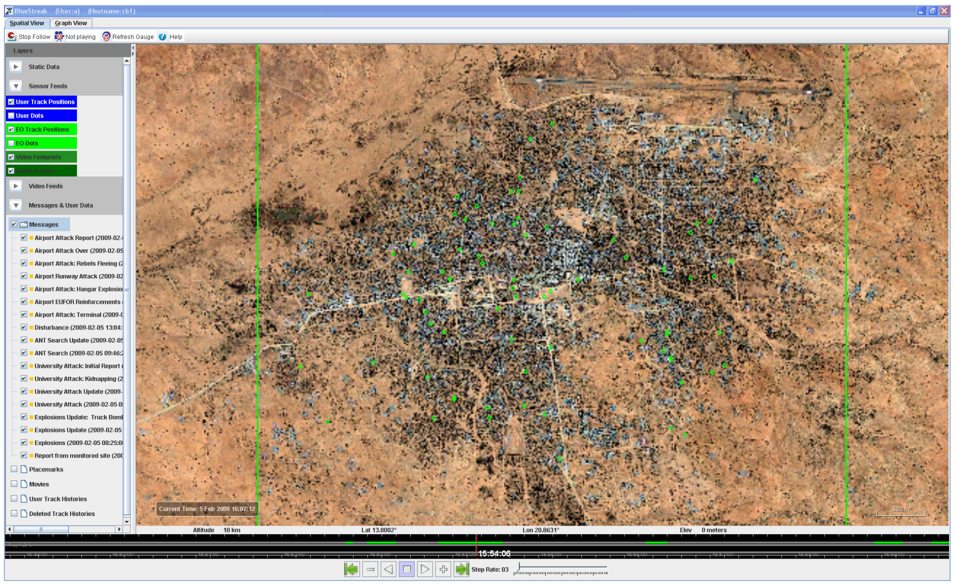


Figure 2.8 *Message panel*

Users can then hover over a message, which reveals a message box, which contains details of the message. The box contains information regarding the message: a report time (when the message was issued), reference time (when the event occurred), details of the message, and the source of the message. An example message box is shown in Figure 2.9.

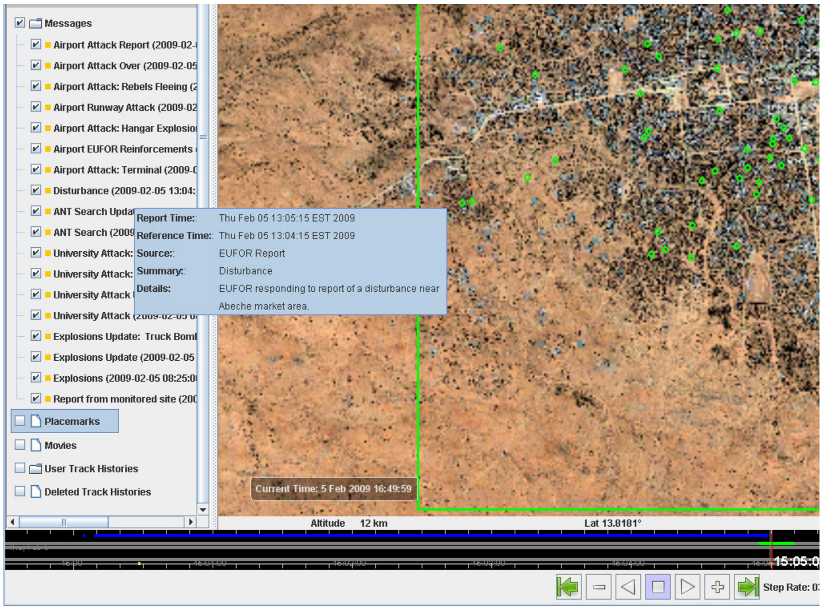


Figure 2.9 *Message details*

After reading the message, if the user decides that it points to a location of interest that the user would like to explore, they can right click on the particular message, which exposes the option to “Fly to.” Upon clicking “Fly to,” the viewport automatically zooms into the location of the message, and centers the timeline to the time of the message.

If the user then wants to denote it as a location of interest (or point of interest, POI), the user can create a “placemark,” by right-clicking at the location and clicking the option to “Make Placemark,” as shown in Figure 2.10.

Figure 2.10 *Creating a "placemark" at location of interest*

A placemark is then made at the location, initially appearing as a yellow dot, as shown in Figure 2.11. A new window appears for the newly created placemark, in which the user enters a placemark name, and labels it with one of six possible colors.

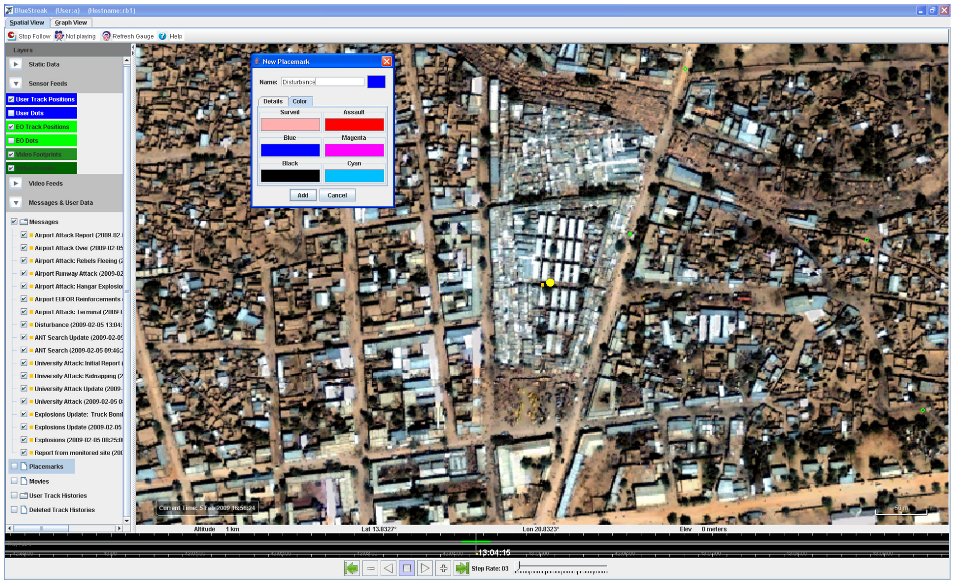


Figure 2.11 *"New Placemark" window*

The color "pink" represents a recommendation by the team to "surveil" the location, while the color "red" represents a recommendation by the team to "assault" the location. Teams are instructed on the difference between the two actions during the mission brief. A recommendation to "surveil" would be upon any suspicious activity, typically transient locations (e.g., meetings spots, observation sites). A recommendation to "assault" would be upon any static red sites, typically buildings (e.g., safe houses, weapons cache).

At the end of the game, teams are only scored on placemarks that are labeled as pink or red. Blue, magenta, black, and cyan are colors that teams can use to organize their placemarks, if they choose to do so, but they are not scored on any placemarks labeled with any of these colors. Typically, at the beginning of the game, teams mark their sites as one of the four innocuous colors, as the suspicion level is very low. As they go through the process and gain more information, their confidence increases. Usually, teams change the colors of the sites they believe to be "red" towards the latter part of the game, to either pink or red. However, these decisions are a continuous process, in that, teams can change the colors of sites at any point in the game.

While teams are given the option to take action against a site in one of two manners, for the purposes of this research, performance is measured simply as the team's ability to discover and label the site as "red," with no distinction being made as to whether they labeled it pink or red.

Once a placemark has been made, the next step is to discover which vehicles have started or stopped at this location of interest. The tool offers a simple way to perform that task, by right clicking on the placemark, and selecting "Make Nomination," as shown in Figure 2.12.

Figure 2.12 *"Create nomination" feature*

A "nomination" is created when a user uses the tool to query the database to display any track stops (start or end) that occurred within a temporal and geographic window that is defined by the user. As such, within the "Nomination" window, shown in Figure 2.13, the user defines a time window and a radius within which there were any track stops. The radius for the nomination is visually displayed on the map as a blue circle.

Figure 2.13 *Nomination window*

Once the user determines the time window and radius and selects “Add,” the tool goes through the database to find all of the track stops within those constraints. The smaller the time window and/or smaller the radius, the user will have fewer tracks to consider, but may possibly miss out on some potential red tracks. The larger the time window and/or larger the radius, the user will have more tracks to filter through, but also ensures that they will not miss out on any potential “red” tracks.

As the system discovers these tracks, it populates the viewport with the tracks, a small blue square representing the track stop location, and the connecting solid blue line representing the track extent, as shown in Figure 2.14. As tracks are added, the track numbers are also displayed in a temporary black box in the upper right hand corner to alert team members that nominated tracks have been added, generated by a member of the team. Each generated track is given a unique track identification number. If the user right clicks on a track stop, information about the track is able to be accessed, such as the track identification number or stop time (the time when the vehicle was stopped at that location). By examining the times of the two track stops of one track, the user can also determine directionality of the track.

Figure 2.14 *Nominated tracks*

As the user views all the tracks, if a depth-first search strategy is employed, then the user can follow one track to its end, creating a placemark and nomination at that end location, and continuing the process in a depth-first approach. Or, the user can follow all the tracks to their ends from the initial point, and then generate placemarks and nominations at each of those ends, and continue the process in a breadth-first approach.

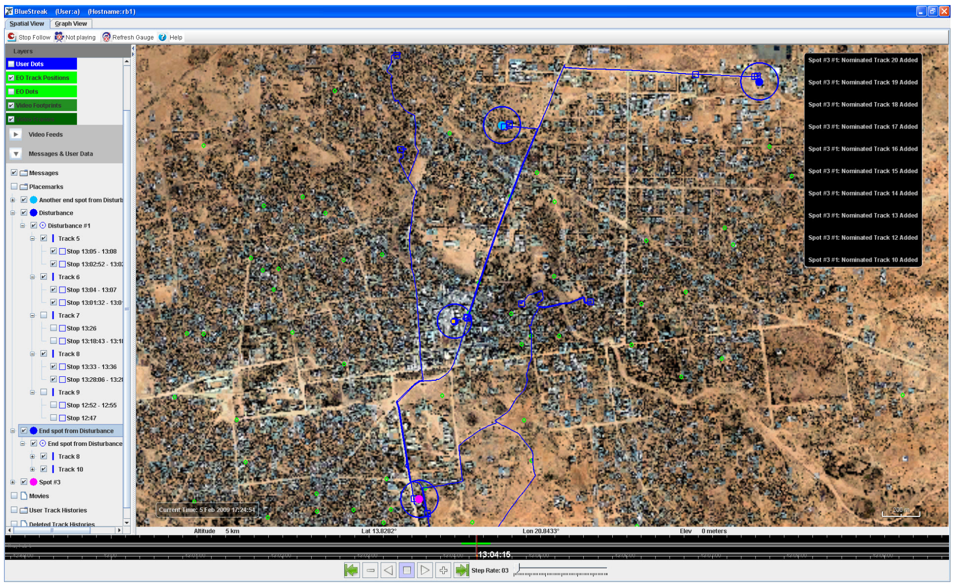


Figure 2.15 *Subsequent placemarks and nominations*

At the end of the game, teams are scored on their ability to identify sites as red locations. Typically, the final decision making process is a team process, in which teams discuss the various options, and decide whether to assault or surveil a particular placemark.

2.3.3 Two-stage decision process

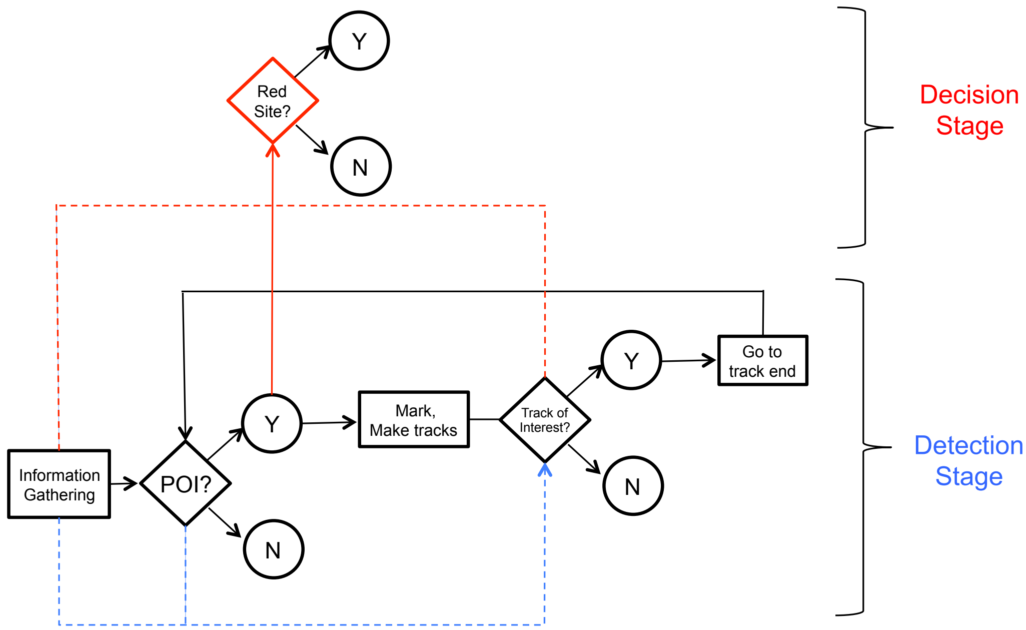


Figure 2.16 *Two stage decision process*

The "Red/Blue" game has been analyzed and decomposed into a two-stage decision process, as shown in Figure 2.16. The first stage is the detection phase, which is when information is gathered and produced, in which teams spend their resources to make placemarks, which are detections of potential targets, or points of interest (POIs). Information gathering primarily takes place through the content within the messages. Information production takes place in the form of finding potential sites and associated tracks. Although initial detections are spurred by information from the given messages, subsequent detections are guided by a secondary task, which is to decide whether there are associated tracks of interest to the POIs. These tracks of interest will lead from one POI to other POIs, and repeats itself in that cycle. Teams can choose to select POIs and tracks of interest in an "open-loop" fashion, and mark as many sites as they can, off of all of the tracks that spawn off of the initial POIs. Or, teams can decide to act in more of a "closed-loop" fashion, by integrating the initially gathered information, as well as previously selected POIs in deciding which tracks are of interest, as denoted by the dotted blue line.

The second stage, the decision stage, is where teams make their decisions about which targets to take action against. This is when teams need to utilize the information that they've gathered during the detection phase and make decisions from it. Effective teamwork will aid this effort, as different aspects of the knowledge acquired during the detection stage may reside with different team members. Thus, effective information sharing and synthesis of the disparate sources of knowledge can lead to effective decision making.

Further details regarding the two stages will be described in Chapter 3.

2.4 Performance Metrics

For both Phase 1 and Phase 2 experiments, the primary measure of performance was the number of hits and false alarms (from the placemarks that the team declares as a red site or red activity), and the corresponding probability of hits and probability of false alarms. Using those probabilities, performance was further quantified by *d'* of SDT, in both the detection phase and decision phases of the exercise.

A questionnaire was given to the participants prior to the exercise regarding demographic information, and a questionnaire was also given immediately after the exercise regarding participants' subjective assessments of their game experience. NASA Task Load Index (NASA-TLX) was the instrument used to measure individual workload (Hart & Staveland, 1988). Situational Awareness Rating Technique (SART) was the instrument used to measure subjective situational awareness (Taylor, 1990). For Phase 1 experiments, additional subjective assessments were posed to the participants. Five questions developed at Aptima, Inc. were used to measure subjective assessments of teamwork. A modified team workload scale (Hildebrand et al., 2003) was used to take gauge how ratings of team workload from the scale would unfold. In addition, questions regarding mutual helpfulness, mutual workload, and role identification were posed in the subjective assessment questionnaire. Lastly, communication interactions were counted as the number of simple utterances a particular team member makes to another member. The interaction counts were used to derive a ratio of communication interactions.

3

Phase 1

Phase 1 experiments were conducted to evaluate the impact of team size and number of resources (people and computers) on teamwork and performance. This phase served to evaluate the bounds of the game, in terms of the number of people and the number of computers per team, which allowed an understanding of the feasibility of the exercise. In addition, these experiments addressed two important questions: (1) what are the characteristics of the performance of one individual with one computer? And, (2) is this an exercise in which having a team does indeed provide a benefit? If so, what factors of a team provide this benefit? Using the individual participant condition as the baseline, the impact of additional people and additional computers could be better understood.

In addition, the performance of one individual could define the basic "taskwork" component of the exercise, in that with only one person, there are no "teamwork" requirements. This would be consistent with the definitions by Salas, et al. (2008) of *taskwork* as "the components of a team member's performance that do not require interdependent interaction with other team members," and *teamwork* as "the interdependent components of performance required to effectively coordinate the performance of multiple individuals." The "Red/Blue" exercise could provide a means of separating those two entities.

Team size research has yielded many disparate results. One may assume that adding more people as resources to solve a problem can only help performance by relieving the load upon each individual, as there are more people to assist in the task. And although it is true that a larger team could enhance performance, having more team members could also hinder decision making effectiveness and performance (Amason & Sapienza, 1997), and several studies have reported performance decrements associated with increases in team size (Amason & Sapienza, 1997; Salas et al., 1999; Horwitz, 2005; Bradner & Mark, 2002) due to the increasing load imparted by increased teamwork requirements.

In the "Red/Blue" exercise, in exploring the effect of team size, Phase 1 results could indicate an "optimal size" (Campion et. al, 1996; Guastello, 2010) that emerges as being appropriate for this exercise. And as team size continues to increase beyond this optimal size, at the other extreme, it could be revealed that teamwork demands, such as coordination and communication demands for larger teams could result in degraded performance (Horwitz, 2005).

As these various perspectives show, there was an exploratory component to Phase 1 in evaluating the team size effect with the "Red/Blue" exercise, which provided important insights that were leveraged for the Phase 2 experiments.

One significant development from Phase 1 was the decomposition of the exercise into a two-stage decision process, as referred to in Chapter 2, Section 2.3.3.

Signal Detection Theory (SDT), as described in Chapter 1, Section 1.7, was the paradigm by which game performance was measured, using the two-stage approach to evaluate progress of performance through the two stages.

3.1 Methodology

3.1.1 Participants

119 men and 49 women between the ages of 18 to 63 ($M = 33.8$, mode = 23, median = 32 years) participated in this phase of experiments. 16 of the participants (12 men and 4 women) participated as 8-person teams, for a total of 2 teams. 72 participants (57 men and 15 women) participated as 6-person teams, for a total of 12 teams. 24 participants (15 men and 9 women) participated as 4-person teams, for a total of 6 teams. 45 participants (30 men and 15 women) participated as 3-person teams, for a total of 15 teams. Lastly, 10 participants (4 men and 6 women) participated individually. Data for the Phase 1 experiments were collected at MIT Lincoln Laboratory and the Human Factors and Applied Cognitive Engineering Lab at Tufts University. 4-person, 6-person, and 8-person teams were generated from targets of opportunity provided by workshops at MIT Lincoln Laboratory. 3-person teams and individual participants

were recruited from the student subject pool at Tufts University. Tufts University students received $10 per hour for their participation.

There were a total of 46 teams, with team sizes ranging from 1 person to 8 people, and corresponding differences in the number of computers provided per team. Table 3.1 provides a detailed summary of the distribution of computers by team size, as well as demographic information. The provided values are averages for the teams in the particular team size. Participants rated their computer experience on a scale of 1-5, with a value of "1" representing "never use a computer" to "5" representing "extremely comfortable with a computer." They also rated how often they play video games on a scale of 1-5, with a value of "1" representing "never," "3" representing "once a week," and "5" representing "daily." Lastly, each participant reported how many years of experience they had in intelligence analysis.

Table 3.1 *Phase 1 summary*

Team Size	Number of Teams	Number of Computers per Team	Age	Computer experience (1-5)	Video game frequency (1-5)	Intelligence analysis experience (years)
1	10	1	23.4	4.2	2.2	0.0
3	15	2	26.0	4.8	2.6	0.6
4	6	3	31.6	4.7	2.0	0.0
6	12	4	39.4	4.3	2.3	1.8
8	2	6	45.2	4.5	1.9	6.7

As Table 3.1 shows, the study of the team size effect began with understanding the characteristics of the performance of an individual (one person with one computer) in this game. 10 individuals, who were Tufts University students, participated in that condition. The sample sizes for the 4-person and 8-person teams were small, as they were the provided targets of opportunity. As such, these teams were not subjected to detailed data analysis, but observed in terms of indications and trends. As shown in Table 3.1, the two 8-person teams had the greatest number of years of intelligence analysis experience. And although it is a small sample size, the results will indicate that the large team size induced teamwork demands that outweighed the value of the intelligence analysis experience in those teams, which potentially resulted in performance drop-offs.

3.1.2 Procedure

All participants followed the general methodology procedure described in Chapter 2 (General Method).

3.2 Results

Table 3.2 summarizes the results for the various team sizes, the framework being the two-stage decision process model presented in

Chapter 2, Section 2.3.3. Full descriptives for all of the variables are displayed in Appendices A-D. The 3-person team descriptives are displayed in Appendix F. Performance, as described by SDT, was measured in terms of the number of hits and false alarms, which were used to calculate probability of hits (*P(H)*) and probability of false alarms (*P(FA)*) at each stage. These values were ultimately used to calculate *d'* at each of the stages.

Table 3.2 *Phase 1 results summary (mean values)*

	Team Size	Place marks	Actions	Hits	FAs	P(H)	P(FA)	d'
Detection	1	12.8		4.6	8.2	0.46	0.04	**1.63**
	3	14.5		5.0	9.5	0.50	0.05	**1.68**
	4	21.8		5.8	16.0	0.58	0.09	**1.61**
	6	20.5		4.8	15.7	0.48	0.08	**1.37**
	8	28.0		5.0	23.0	0.50	0.12	**1.25**
Decision	1		6.9	3.5	3.4	0.74	0.42	**0.96**
	3		5.5	3.5	2.0	0.67	0.22	**1.37**
	4		9.0	3.7	5.3	0.64	0.34	**0.81**
	6		7.3	3.3	4.0	0.68	0.24	**1.37**
	8		12.0	3.5	8.5	0.71	0.34	**0.96**
Total	1			3.5	3.4	0.35	0.02	**1.75**
	3			3.5	2.0	0.35	0.01	**1.97**
	4			3.7	5.3	0.37	0.03	**1.58**
	6			3.3	4.0	0.33	0.02	**1.75**
	8			3.5	8.5	0.35	0.05	**1.41**

Table 3.2 shows the average values for each metric in the column for each team size. The number of "Placemarks" describes the average number of

detections teams made at the detection stage. These are marks placed on points of interest, as the teams have not yet decided whether they are hostile. The number of "Hits" at the detection stage shows the average number of correct red sites they have marked out of their total number of placemarks. Likewise, the number of "False Alarms (FAs)" shows the average number of non-red sites they have marked as detections. The probability of hits ($P(H)$) is the number of hits divided by 10, which is the total number of red sites there are in the game.

$$P(H)_{det} = \frac{H_{det}}{10}$$

The probability of false alarms ($P(FA)$) is the number of FAs divided by 190, which is the total number of non-red sites in the game.

$$P(FA)_{det} = \frac{FA_{det}}{190}$$

Detection stage performance, or the team's sensitivity (d') at the detection stage (d'_{det}) is expressed in standard deviation units, and is calculated as the difference between the normalized means (z scores) of $P(H)_{det}$ and $P(FA)_{det}$.

$$d'_{det} = z[P(H)_{det}] - z[P(FA)_{det}]$$

From Table 3.2, during the decision stage, the number of "actions" describes the average number of placemarks that teams decided is hostile, and decided to take action on. The number of "Hits" at the decision stage is the average number of correct red sites they have marked as hostile. Likewise, the number of "False Alarms (FAs)" is the average number of non-red sites they have marked to take action on. The probability of hits ($P(H)$) is the number of hits divided by the number of red sites the team had discovered during the detection stage, which could range from 0 to 10.

$$P(H)_{dec} = \frac{H_{dec}}{H_{det}}$$

The probability of false alarms ($P(FA)$) is the number of FAs divided by the number of FAs the team had detected during the detection stage.

$$P(FA)_{dec} = \frac{FA_{dec}}{FA_{det}}$$

Decision stage performance, or team sensitivity (d') at the decision stage (d'_{dec}) is expressed in standard deviation units, and is calculated as the difference between the normalized means (z scores) of $P(H)_{dec}$ and $P(FA)_{dec}$.

$$d'_{dec} = z[P(H)_{dec}] - z[P(FA)_{dec}]$$

Lastly, in describing overall performance, from Table 3.2, the total (or overall) number of "Hits" is the average number of correct red sites teams have marked as hostile (to take action against). Likewise, the number of "False Alarms (FAs)" is the average number of non-red sites they have marked to take action against. The total (or overall) probability of hits (*P(H)*) is the number of hits divided by 10, which is the total number of red sites.

$$P(H)_{tot} = \frac{H_{dec}}{10}$$

The probability of false alarms (*P(FA)*) is the number of FAs divided by 190, which is the total number of non-red sites.

$$P(FA)_{tot} = \frac{FA_{dec}}{190}$$

Overall performance, or total (overall) team sensitivity (*d'*) at the decision stage (d'_{tot}) is expressed in standard deviation units, and is calculated as the difference between the normalized means (*z* scores) of $P(H)_{tot}$ and $P(FA)_{tot}$.

$$d'_{tot} = z[P(H)_{tot}] - z[P(FA)_{tot}]$$

Figure 3.1 presents the results shown in Table 3.2, the values for the probability of hits (*P(H)*) and false alarms (*P(FA)*), as a traditional Receiver Operator Characteristic (ROC) plot, which is a graphical representation of the point estimate of the sensitivity of the various team sizes at the detection and decision stages, and at the overall level.

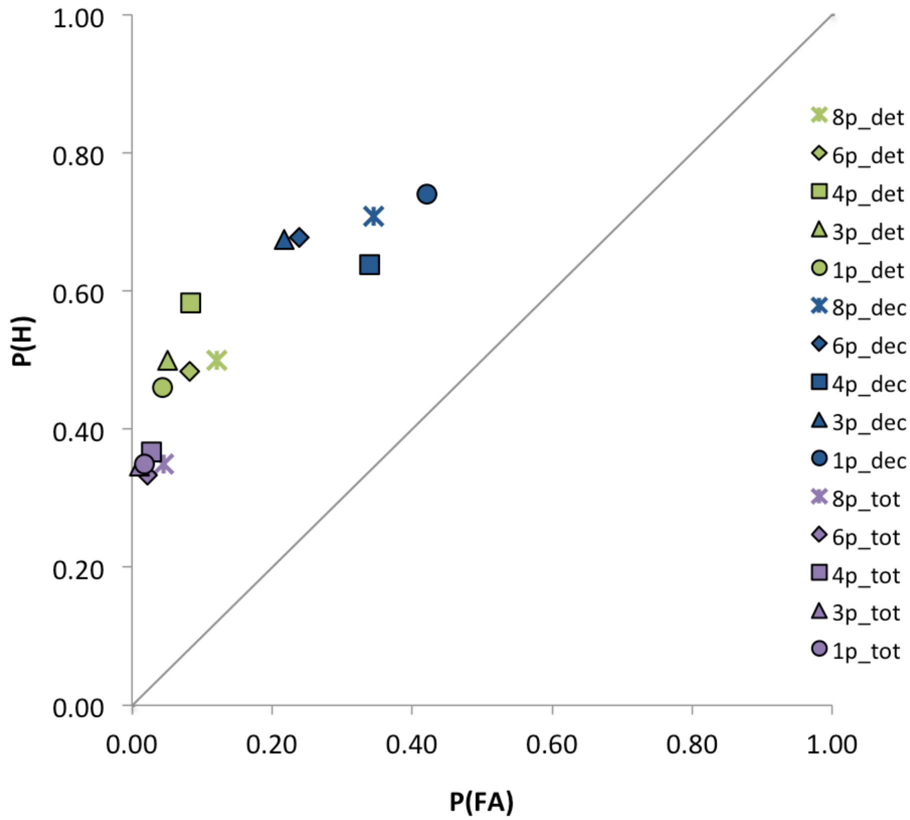


Figure 3.1 *Phase 1 results on ROC graph*

Regarding notation for the figures in Chapter 3, "8p" refers to the 8-person teams, "6p" refers to the 6-person teams, "4p" refers to the 4-person teams, "3p" refers to the 3-person teams, and "1p" refers to the "1-person team," or individual participants. The labels next to each of the team designations refer to the stage: "det" is detection stage, "dec" is decision stage, and "tot" is the total or overall result.

Analysis of the data was performed primarily through one-way independent measures ANOVA to uncover statistically significant differences in the data. When the assumptions of normality or homogeneity of variances were not met, a nonparametric test, Kruskal-Wallis one-way analysis of variance, was performed. The Shapiro-Wilk test was used to test for normality, and Levene's test was used to test for homogeneity of variances. When statistically significant results were discovered, the Tukey HSD or Dunn test was used to perform post-hoc paired comparisons.

3.2.1 Detection Stage, Decision Stage, Overall Sensitivity (*d'*)

Overall detections (placemarks) and number of decisions (actions) teams make at the two stages were measured and used to calculate the probability of hits ($P(H)$) and probability of false alarms ($P(FA)$), which in turn were used to calculate sensitivity (d'). Some insights can be made in examining the differences in d' values at the various stages. First of all, sensitivity at the detection stage (d'_{det}) was similar for all teams, which could indicate that in general, teams have comparable detection performance. A nonparametric test, the Kruskal-Wallis test, showed that d'_{det} was not statistically different across team size, $\chi^2(4, N=45) = 4.88$, $p = .300$. The mean d'_{det} for 1-person teams was 1.63 (SD = .65), the mean

d'_{det} for 3-person teams was 1.68 (SD = .45), the mean d'_{det} for 4-person teams was 1.61 (SD = .22), the mean d'_{det} for 6-person teams was 1.37 (SD = .30), and the mean d'_{det} for 8-person teams was 1.25 (SD = .16).

Figure 3.2 depicts the average *d'* values for the different stages by team size that illustrates the relatively comparable detection stage performance, as well as the relatively differing decision stage performance, particularly between the individual participants and the 3-person and 6-person teams.

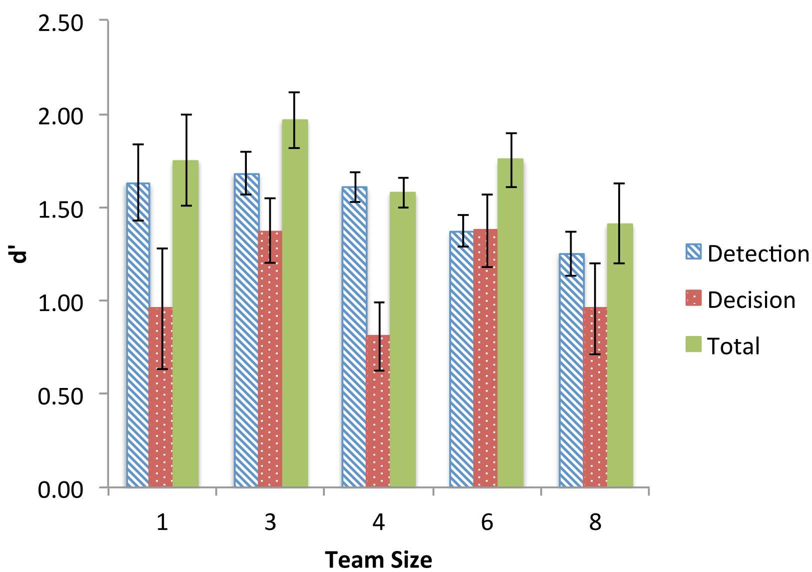


Figure 3.2 *Performance (d' values) by team size (mean values)*

The decision stage is where teams make their decisions about which targets to take action against. A differentiation occurs at this stage, between teams that effectively utilize the information generated during the detection phase, and teams that do not effectively utilize this information. The task of detecting potential targets is a relatively straightforward task. However, the decision stage is more complicated, and involves synthesizing and using the knowledge that was acquired during the detection stage. This usage of the produced information from the detection stage involves teamwork, as team members need to share, organize, and coordinate the information in order to make team decisions about which sites to take action against. As was the case with d'_{dec}, a simple one-way independent measures ANOVA shows that for decision stage sensitivity (d'_{dec}), there was no significant main effect for team size, $F(4, 40) = 1.10$, $p = .368$, and the Kruskal-Wallis test showed that for d'_{tot}, there was no significant main effect for team size, $\chi^2(4, N=44) = 3.69$, $p = .449$. However, the figure of merit is not so much in the values of d'_{dec} and d'_{tot} themselves, and simply examining the results based on team size is not sufficient.

Rather, of more interest is the understanding of how teams utilize the information generated during the detection phase in their decision phase, and how team size plays a factor in that process. And so the metric of concern becomes the change in sensitivity from detection to decision, and from detection to overall sensitivity.

3.2.2 Delta decision (Δ_{dec}) and Delta total (Δ_{tot})

Using these observations and insights, SDT and the two-stage decision process framework were used to derive two metrics, Δ_{dec} and Δ_{tot}. These two metrics are presented as emergent properties of teamwork. Δ_{dec} is defined as the difference between a team's *d'* at the decision stage (d'_{dec}) and its *d'* at the detection stage (d'_{det}).

$$\Delta_{dec} = d'_{dec} - d'_{det}$$

Δ_{tot} is defined as the difference between a team's total (overall) *d'* (d'_{tot}) and its *d'* at the detection stage (d'_{det}).

$$\Delta_{tot} = d'_{tot} - d'_{det}$$

These metrics can be used to describe how well teams are utilizing their detection stage information in their decision making, and in turn, their overall performance. During the detection stage, teams gather as much information from which they will have to make their decision. For example, they will not accrue any more "Hits" than they have at the detection stage. As such, teams that evaluate this produced information effectively, by holding onto as many of the hits as possible, and reducing as many of the false alarms as possible, will in turn have high Δ_{dec}.

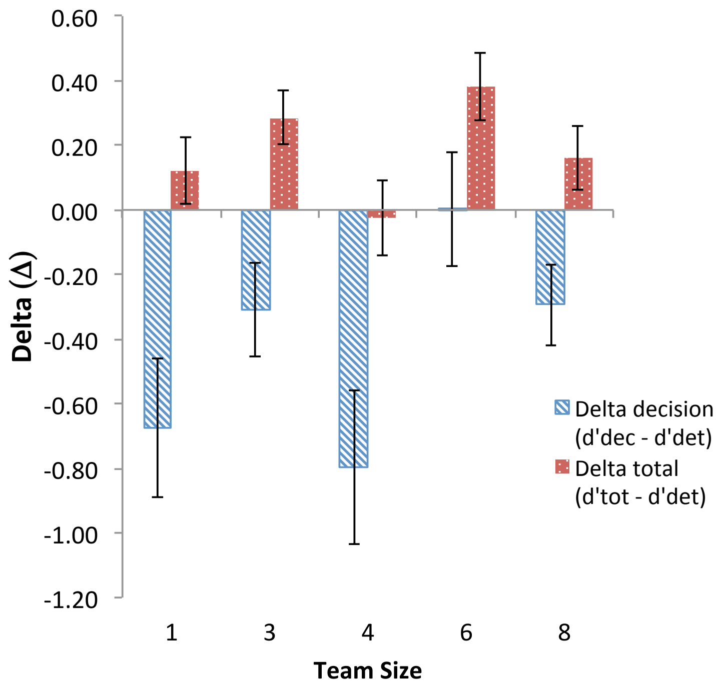


Figure 3.3 *Delta decision (Δ_{dec}) and Delta total (Δ_{tot}) by team size (mean values)*

A plot of Δ_{dec} and Δ_{tot} by team size is shown in Figure 3.3. The plot shows indications of a parabolic effect of team size. Both Δ_{dec} and Δ_{tot} tend to improve with increase of team size, but there are indications of a drop-off with the 8-person teams. As previously mentioned, 4-person and 8-person teams had small sample sizes, but the results were used to observe trends. One-way independent measures ANOVA showed that for Δ_{dec} and Δ_{tot}, there were strong indications of main effects for team size [$F(4, 40) = 2.53$, $p = .055$; $F(4, 40) = 2.20$, $p = .087$], showing that these derived metrics are discernible across team size, and can potentially be used to characterize the effect team size has on the change in sensitivity from detection to decision, and detection to overall performance. The mean Δ_{dec} for 1-person teams was -.67 (SD = .68), the mean Δ_{dec} for 3-person teams was -.31 (SD = .56), the mean Δ_{dec} for 4-person teams was -.80 (SD = .63), the mean Δ_{dec} for 6-person teams was .002 (SD = .61), and the mean Δ_{dec} for 8-person teams was -.29 (SD = .18). The mean Δ_{tot} for 1-person teams was .12 (SD = .32), the mean Δ_{tot} for 3-person teams was .29 (SD = .32), the mean Δ_{tot} for 4-person teams was -.03 (SD = .31), the mean Δ_{tot} for 6-person teams was .38 (SD = .36), and the mean Δ_{tot} for 8-person teams was .16 (SD = .14).

3.2.2.1 Team Categories

The results for all of the teams organized by team size and based on Δ_{dec} and Δ_{tot}, as shown in Figures 3.4 and 3.5, showed some emergent groupings of teams. These grouping categories are proposed as descriptions of how that particular category team may be using the knowledge acquired during the detection stage towards their decision making and overall performance.

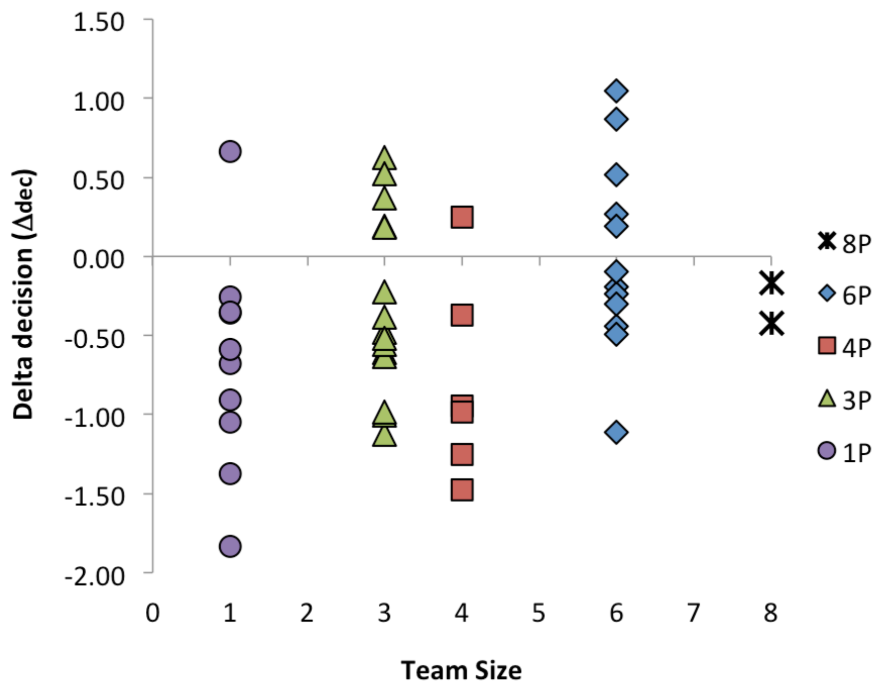


Figure 3.4 *Delta decision (Δ_{dec}) by team size*

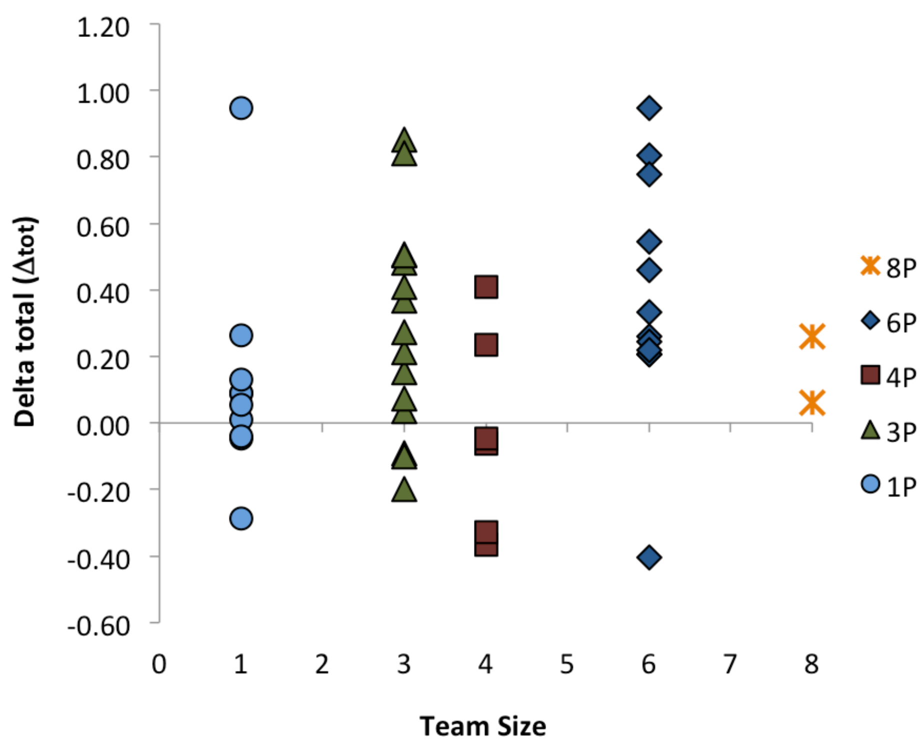


Figure 3.5 *Delta total (Δ_{tot}) by team size*

Three categories are proposed, based on Figures 3.4 and 3.5:

Category 1: $\Delta_{dec} > 0$

- Indicates effective use of detection stage knowledge in decision making

Category 2: $\Delta_{dec} < 0$ & $\Delta_{tot} > 0$

- Indicates less effective use of detection stage knowledge in decision making, but still able to use their decision stage performance in a way to achieve improved overall performance from detection stage performance

Category 3: $\Delta_{dec} < 0$ & $\Delta_{tot} < 0$

- Indicates ineffective use of detection stage knowledge in decision making, and resulted in poor overall performance

Of the 46 teams from Phase 1, there were 12 teams in Category 1, 23 teams in Category 2, and 11 teams in Category 3. Table 3.4 and Figure 3.6 summarize the results for sensitivity (d') and delta (Δ) values for the different stages, based on the proposed categories.

Table 3.3 *Phase 1 results by team categories (mean values)*

Category	Number of teams	d'_{det}	d'_{dec}	d'_{tot}	Δ_{dec}	Δ_{tot}
1	12	1.52	1.99	2.19	0.47	0.67
2	22	1.50	1.01	1.69	-0.49	0.22
3	11	1.72	0.65	1.54	-1.07	-0.18

A one-way independent measures ANOVA showed that there was a significant main effect for team category on Δ_{dec}, $F(2, 42) = 68.30$, $p = .001$. Tukey's HSD tests showed that each of the three categories had statistically significant differences to one other. The mean Δ_{dec} for Category 1 was .47 (SD = .28), the mean Δ_{dec} for Category 2 was -.49 (SD = .30), and the mean Δ_{dec} for Category 3 was -1.07 (SD = .39). Likewise, the Kruskal-Wallis test showed that there was a significant main effect for team category on Δ_{tot}, χ^2 $(4, N=45) = 36.63$, $p = .001$. Post-hoc tests showed that all of the pairwise comparisons between each of the categories were statistically significant. The mean Δ_{tot} for Category 1 teams was .67 (SD = .20), the mean Δ_{tot} for Category 2 teams was .22 (SD = .13), and the mean Δ_{tot} for Category 3 teams was -.18 (SD = .14).

A one-way independent measures ANOVA showed that there was no significant main effect for team category on d'_{det}, $F(2, 42) = .95$, $p = .394$. The mean d'_{det} for Category 1 was 1.52 (SD = .35), the mean d'_{det} for

Category 2 was 1.50 (SD = .50), and the mean d'_{det} for Category 3 was 1.72 (SD = .42).

These results were consistent with the ANOVA results for team size, and further supported the observations of the two stage process. As mentioned, the task of detecting potential targets is a relatively straightforward task. But the differentiation amongst teams occurred at the decision stage, in how teams utilized the information gathered and produced at the detection stage. For d'_{dec}, a one-way independent measures ANOVA showed that there was a significant main effect for team category, $F(2, 42) = 18.91$, $p = .001$. Tukey's HSD tests showed that d'_{dec} for Category 1 teams was statistically significantly higher than both Category 2 and 3 teams. Category 2 teams did not differ significantly from Category 3 teams. The mean d'_{dec} for Category 1 was 1.99 (SD = .51), the mean d'_{dec} for Category 2 was 1.01 (SD = .59), and the mean d'_{dec} for Category 3 was .65 (SD = .52). Likewise, for d'_{tot}, a one-way independent measures ANOVA showed that there was a significant main effect for team category, $F(2, 42) = 5.10$, $p = .010$. Tukey's HSD tests showed that d'_{tot} for Category 1 teams was statistically significantly higher than both Category 2 and 3 teams. Category 2 teams did not differ significantly from

Category 3 teams. The mean d'_{tot} for Category 1 was 2.19 (SD = .51), the mean d'_{tot} for Category 2 was 1.69 (SD = .56), and the mean d'_{tot} for Category 3 was 1.54 (SD = .47). These results show that the three categories can indeed be used to distinguish team performance at the decision and overall phases. The d' values for the three categories are shown in Figure 3.6.

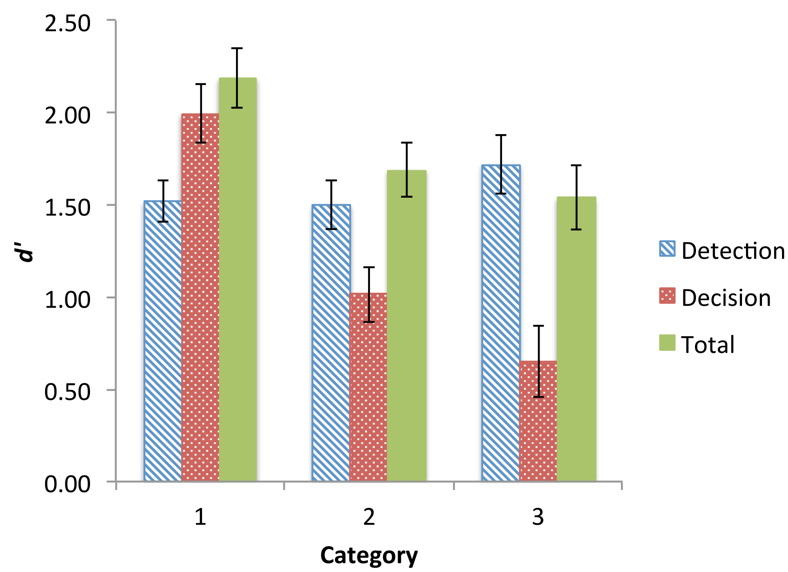


Figure 3.6 *d' values by team categories (mean values)*

As described by the ANOVA results, and as seen in Figure 3.6, d'_{det} does not differ across the categories, but the difference in d'_{dec} is clearly seen, as well as the difference in overall performance across the three categories.

3.2.3 Impact of Resources

In addition to examining the results based on the categories, further investigation of the difference in *d'* between the two stages revealed that increasing resources (people and computers) did not significantly increase the number of correct actions (hits). Rather, increasing resources resulted in a statistically significant increase in the number of false alarms, as shown in Figures 3.7 and 3.8.

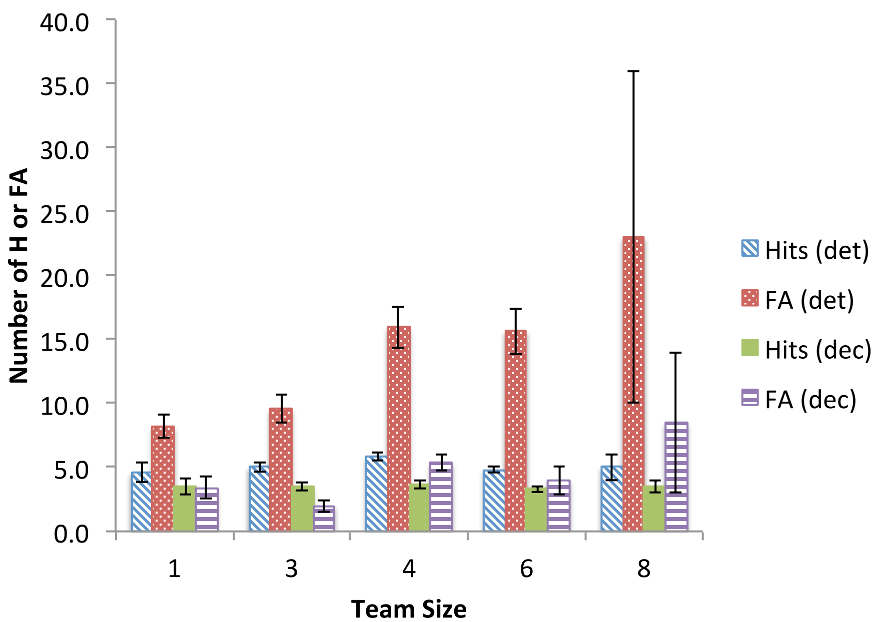


Figure 3.7 *Number of Hits (H) or False Alarms (FA) by team size (mean values)*

The large SE for the 8-person teams was due to the small sample size. The Kruskal-Wallis nonparametric test showed that the number of hits at the detection stage was not statistically different across team size, $\chi^2(4, N=45)$

$= 2.93$, $p = .569$, and the number of hits at the decision stage was not statistically different across team size, $\chi^2(4, N{=}45) = 0.436$, $p = .979$. However, the Kruskal-Wallis test showed that the number of false alarms at the detection stage was statistically different across team size, $\chi^2(4, N{=}45) = 17.06$, $p = .002$, and the number of false alarms at the decision stage had strong indications of statistically significant differences across team size, $\chi^2(4, N{=}45) = 8.74$, $p = .068$. And as Figure 3.7 further indicates, this showed that increasing resources had the effect of potentially bringing more noise into the process, in the form of more placemarks. Team size and the total number of placemarks were strongly correlated, $r(43) = .545$, $p = .001$. Team size and false alarms at the detection stage were also strongly correlated, $r(43) = .563$, $p = .001$. Team size and false alarms at the decision stage had strong indication of correlation, $r(43) = .272$, $p = .070$.

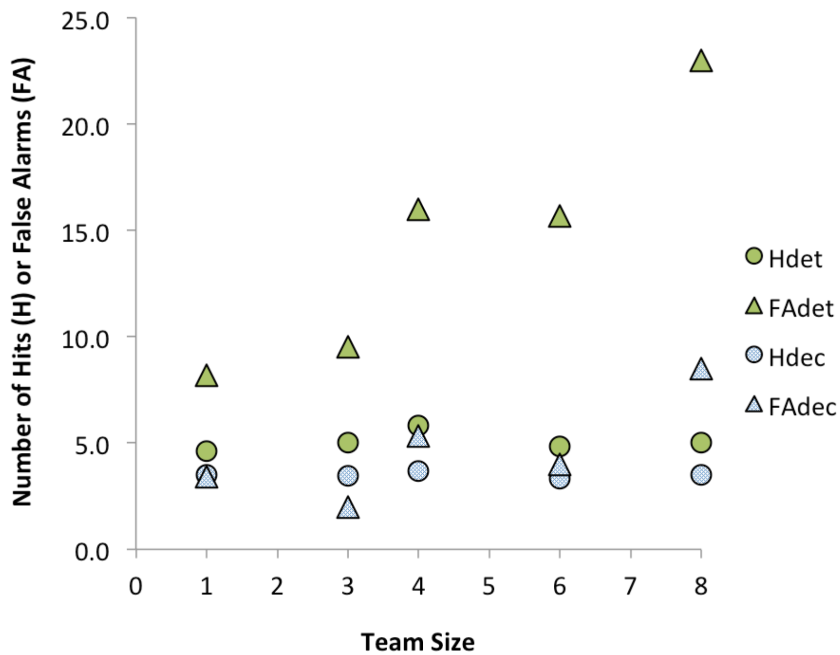


Figure 3.8 *Number of Hits (H) or False Alarms (FA) by team size (mean values)*

As Figure 3.8 shows, increasing the number of computers and the number of people increased the noise that teams had to filter through. And as Figure 3.9 indicates, effective performance may have been driven more by a team's ability to reduce false alarms than its ability to find more red sites. All teams, as shown in Figure 3.8, discovered approximately the same number of potential hits at the detection stage, as there was not a significant difference. But Category 1 teams distinguished themselves by reducing their false alarms from the detection stage, while holding onto as many as they could of the red sites they found during the detection stage.

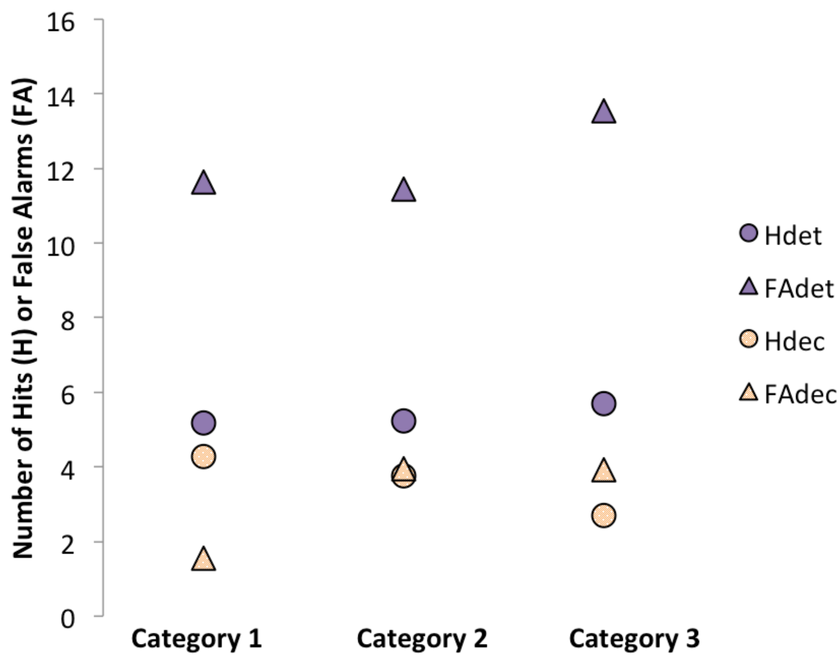


Figure 3.9 *Number of Hits (H) or False Alarms (FA) by team category (mean values)*

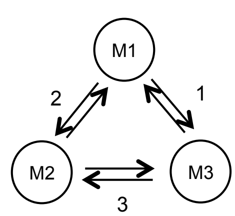
In Figure 3.9, "Hdet" is the number of red sites they found during the detection stage, and "FAdet" is the number of non-red sites they found during the detections stage. "Hdec" is the number of red sites upon which they decided to take action, and "FAdec" is the number of non-red sites upon which they decided to take action.

3.2.4 Communication Interactions

It can be hypothesized that the filtering process described above is related to teamwork and a team's communication interactions. During the game, knowledge regarding sites resides with different team members, and so balanced communication interactions could produce the result of better team filtering of false alarms.

A simple ratio, called the Communication Interaction Ratio (CIR), was used to quantify this balance for teams of three members. The ability to measure communication interactions for the other size teams was not available.

The ratio (CIR) is described below in Figure 3.10. M1, M2, and M3 represent each team member, and the numbers 1, 2, and 3 represent the interaction, or simple utterance, in that link between the particular team members. In calculating CIR, the sum of all the interactions in each link is used, as shown in the equation below. As such, a ratio close to the value of "1" will be indicative of balanced interactions, and values greater than "1" will be indicative of unbalanced interactions.



$$CIR = \frac{\frac{\sum 1 + \sum 2}{\sum 3} + \frac{\sum 1 + \sum 3}{\sum 2} + \frac{\sum 2 + \sum 3}{\sum 1}}{6}$$

Figure 3.10 *Communication Interaction Ratio (CIR)*

Among the 3-person teams that were measured, there was a significant correlation between Δ_{dec} and CIR, $r(6) = -.661$, $p = .051$, which is shown in Figure 3.11.

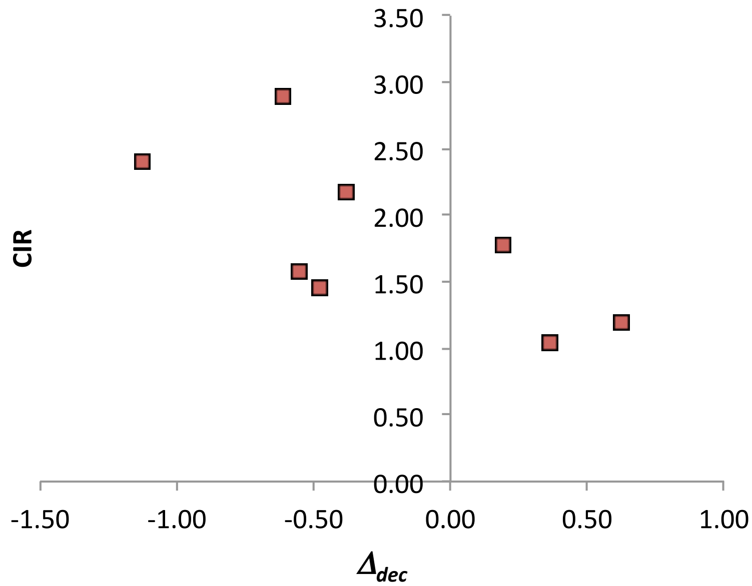


Figure 3.11 *Communication Interaction Ratio (CIR) by Delta decision (Δ_{dec})*

3.2.5 Subjective Assessments

As mentioned in Chapter 2, there were a variety of questions posed to participants regarding their subjective assessments of workload, SA, team workload, teamwork, and mutual helpfulness. The questionnaires that were used for these assessments are provided in Appendix L. For Phase 1, the questionnaires were used in an exploratory manner, to evaluate whether these particular questions had the sensitivity to measure the parameters of the "Red/Blue" exercise. The results for team workload, teamwork, and mutual helpfulness based on these particular questionnaires did not yield any significant observations, and perhaps did not have adequate sensitivity for this exercise. Therefore, these questions were not subject to in-depth analysis, and were not part of the Phase 2 experiments.

However, the results from the subjective questionnaires for workload (NASA-TLX) and situational awareness (SART) provided some additional insight into the process, and are presented below. The Kruskal-Wallis test showed that workload (NASA-TLX) was not statistically different across team size, $\chi^2(4, N=44) = 2.36$, $p = .670$, as shown in Figure 3.12.

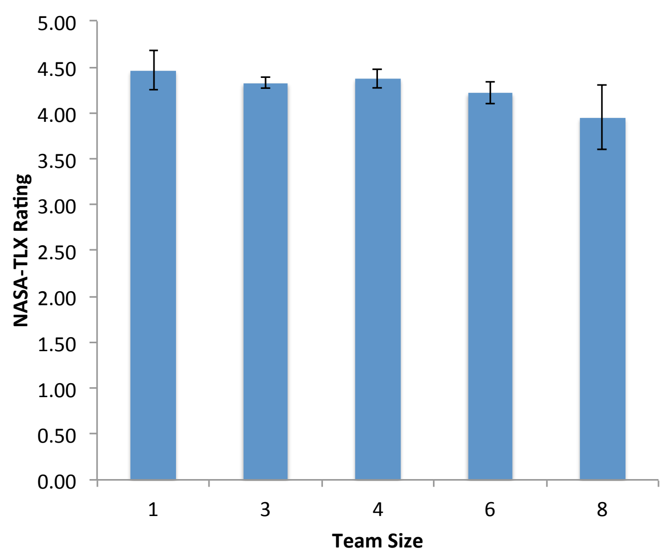


Figure 3.12 *NASA-TLX ratings by team size (mean values)*

The mean workload for 1-person teams was 4.46 (SD = .68), the mean workload for 3-person teams was 4.33 (SD = .32), the mean workload for 4-person teams was 4.37 (SD = .28), the mean workload for 6-person teams was 4.22 (SD = .42), and the mean workload for 8-person teams was 3.95 (SD = .50).

These results indicated that the addition of resources did not alleviate the average individual workload that participants experienced. As this is an aggregate result, it may indicate that since the taskwork component is fairly consistent across teams, and there were always more people with computers than those without, the workload associated with the taskwork was what impacted the non-significant workload result. Or another potential explanation could be that the load induced by additional people and computers may have offset whatever benefit they may have brought to the team, thus resulting in relatively even workload across all team sizes. These are conjectures, and would require more sensitive measurement to conclusively understand this outcome, and are recommended for future work.

For SA, the Kruskal-Wallis test showed that subjective situational awareness ratings (SART) were statistically significantly different across team size, $\chi^2(4, N=45) = 9.55$, $p = .049$, as shown in Figure 3.13.

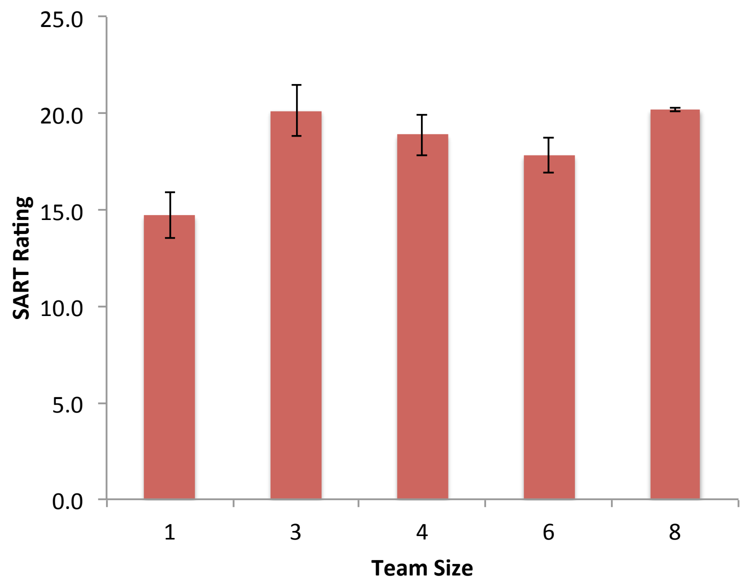


Figure 3.13 *SART ratings by team size (mean values)*

Post-hoc tests showed that the 3-person team SA scores were statistically significantly higher than the 1-person SA scores. But the SA scores between all the other pairwise comparisons were not statistically significant. The mean SA for 1-person teams was 14.7 (SD = 3.74), the mean SA for 3-person teams was 20.2 (SD = 5.25), the mean SA for 4-person teams was 18.9 (SD = 2.83), the mean SA for 6-person teams was 17.8 (SD = 3.06), and the mean SA for 8-person teams was 20.2 (SD = .14).

The results may indicate that having additional people, and perhaps the presence of people without a computer could have resulted in higher SA than individual participants. Anecdotally, those without a computer, in some cases, acted as a facilitator. These people typically utilized the whiteboard or large physical map, as well as facilitated communication amongst team members. These activities could have potentially allowed their team to achieve greater awareness of the situation. This is in contrast to an individual participants' awareness of the situation, which was primarily attained only via the computer (although they were free to use any of the other tools provided to them, such as the whiteboard or map). Again, the factors that lead to the attainment of SA was not a primary focus of this research, but these conjectures could be investigated in future work.

3.3 Discussion

The Phase 1 experiments gave rise to multiple observations, and a convergence of many ideas and theories were able to been seen. These observations were used in the design of the next phase, Phase 2 experiments, designed to further investigate the effect of teamwork (as promoted by resource allocation) on team performance.

Results indicated that a component of teamwork and effective team performance was manifested in the reduction of false alarms. Good team performance was not necessarily about how many hits or targets a team would find, but was actually more directed by the effectiveness of its filtering potential false alarms. This difference can be understood by the ratio of signal to noise (hits to false alarms), best captured by d', a team's sensitivity. Good filtering is a salient feature of a team with good teamwork, which is captured by this sensitivity measure, and that their sensitivity improves in their decision making stage from their detection phase.

Parasuraman, Sheridan, and Wickens (2000) proposed a four-stage model of a typical process: information acquisition, information analysis, decision selection, and action implementation. The “Red/Blue” exercise and its two-stage decision process mimic this model. Information acquisition describes the detection phase. Information analysis and decision selection describes the decision phase. And action implementation describes the final actions that are selected per site.

The results from Phase 1 showed no significant differences in d'_{det} across all teams, regardless of team size or team category. While it cannot be dismissed that differences may indeed exist, the findings support an understanding of d'_{det} as representing taskwork, performing the task of finding points of interests and making placemarks. Therefore, task performance and teamwork emerged as separable entities.

As teams distinguished themselves and showed significant differences in performance improvements from the detection to decision stage, the metrics Δ_{dec} and Δ_{tot} were proposed as properties of teamwork that capture how well teams were using their gathered and produced information in making decisions.

The individual participant results revealed that an individual with one computer is able to do well (d'_{tot}) in the exercise. This demonstrated that the "Red/Blue" exercise is not resource limited, in that one person and one computer are sufficient to achieve good performance.

However, the individual participant results also showed the widest range (2.2) and variance (.60), showing that while an individual was able to do as well as, or in some cases, better than teams of multiple people, individuals were also capable of doing worse than teams of multiple people. Therefore, teams were able to offer the opportunity for greater consistency of performance.

All individual participants had negative Δ_{dec} values, apart from one individual. This gives further support that Δ_{dec} may indeed be a property of teamwork, requiring more than one person to achieve it. A high Δ_{dec} showed that the aggregate had improved knowledge in some way, while many underachieving teams have results that show that they could be no different than individuals, perhaps reflecting a lack of teamwork. So while high Δ_{dec} shows improvement of knowledge, as Δ_{dec} values become more negative, it may reflect negative teamwork in counter productive relationships amongst team members.

The value of d'_{det} reflects the sensitivity that is established at the information gathering stage. It will be impossible for a team to find any more correct red sites beyond the total number found at the detection stage. If Δ_{dec} is low, then the information or sensitivity that was available at the detection stage was not effectively passed through to the decision making process. This would reflect some defect in the teamwork of the team. For higher levels of Δ_{dec}, the indication is that there was sensitivity added at the decision making stage, due to teamwork.

In addition, higher levels of Δ_{dec} increased the likelihood that a team would have improved overall performance relative to its performance at the detection stage. The results showed that regardless of a team's taskwork, as captured by d'_{det}, improved teamwork (high Δ_{dec}) can result in improvement in performance upon its taskwork performance (high Δ_{tot}). This is demonstrated by the strong correlation between Δ_{tot} and Δ_{dec}, $r(43) = .922$, $p = .001$, shown in Figure 3.14.

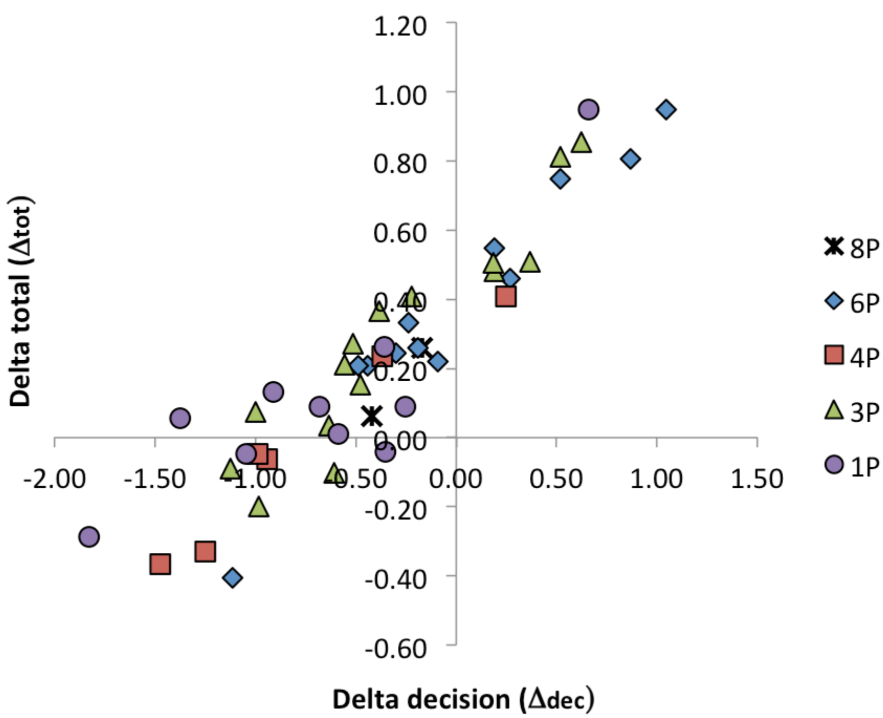


Figure 3.14 *Correlation between Delta total (Δ_{tot}) and Delta decision (Δ_{dec}) in Phase 1*

There is also a significant correlation between team size and Δ_{dec} , $r(43) = .306$, $p=.041$. Figure 3.14 also shows that this correlation is driven primarily by the improvements in Δ_{dec} shown by the 3-person and 6-person teams.

Phase 1 results indicated answers to the question of what value additional resources bring. At the detection stage (information acquisition), their contribution may simply be the addition of more potential placemarks. But at the decision stage (analysis and decision making), they can help filter out the noise. The teams with poorer performances were teams that failed to provide appropriate filtering. The higher performing teams would be those that have the most complete shared mental picture (e.g., effective data processing, filtering). Team members may be found to only help if they can find a way to reduce the noise at whatever level they work at, in addition to finding hits. This could be reflected in the higher SART scores that are seen in the 3-person teams from the 1-person team. This could also be reflected in the higher delta values that were observed as the number of resources increased. Although the number of people increased, the number of computers increased as well, allowing more sources of information and data that may have contributed to a better shared mental picture.

3.3.1 Information Production/Information Synthesis

As a resource, the computer is a source of data allowing a person to generate or produce information. During the Phase 1 experiments, while teams were free to self-organize, two general roles, as related to their interaction with information, were anecdotally observed to emerge as a function of resource. The people who had a computer tended to act in an "analyst" role. People in this role mainly interacted with the computer that was directly in front of them, and performed the task of producing information. People who did not have a computer tended to act in a "facilitator" role, as they did not have their own information source with which to interact. Without their own computer, people in this role primarily interacted with the whiteboard, map, or projected view of an analyst's screen. It is conjectured that without their own information source, they adopted other tools that were made available to them in order to contribute to the team's information processing. In that sense, they were observed, using those other tools, performing the task of organizing and synthesizing the information being produced by the analysts. They were also observed performing a variety of other facilitation or leadership tasks, including distributing tasks to individual analysts, facilitating communication between analysts, and maintaining the "big picture" perspective.

3.3.2 Conclusions

In Phase 1, the number of resources were indicated to be beneficial when used to properly filter false alarms, lowering the team's signal to noise ratio. These concepts are best captured by SDT's *d'* metric, which captures a team's sensitivity. The two-stage approach allowed the decomposition of the game into a detection and decision stage. This approach also allowed for the understanding of teamwork as one in which teams effectively use their produced and synthesized information to make accurate decisions, allowing for an overall improvement in performance relative to their detection stage performance. The metrics Δ_{dec} and Δ_{tot} captured the contribution of teamwork to performance.

Phase 1 allowed an understanding of the contributions of the two stages (d'_{det}, d'_{dec}) to a team's overall performance (d'_{tot}). As the concern in most cases regards how to achieve high performance, Phase 1 regression results show that both higher d'_{det} scores and/or higher d'_{dec} scores translate to higher d'_{tot}, with both having their own unique contribution to overall performance. As the two-stage decision model has proposed the detection stage representing the taskwork of the game, efforts can be made to improve d'_{det} by improving the task process. A variety of conjectures could be made, such as improved GUI design, larger screen displays, or different training regimen, to name a few. However, as the focus of this research is teamwork, and as the two-stage process allows a method to segregate taskwork from teamwork, the focus for the Phase 2 experiments was on the decision stage, and how through teamwork, teams can improve upon its taskwork in this stage to improve its overall performance.

Results showed that one person and one computer are sufficient conditions for good performance in this game. However, the results also showed that teams (adding people) have the potential to give greater consistency to performance. This consistency was the result of teamwork, which resulted from sharing of information and balanced communication interactions. In short, adding more people and resources can increase the consistency of high performance in the game, but it doesn't always result in that way, in that it is not simply about putting more eyes on the problem.

Increasing the number of people on a team is only beneficial if the additional people are able to contribute to increasing Δ_{dec} and Δ_{tot}. Once team size grows beyond the ability for people to contribute, potentially due to increased coordination demands, then poorer performance could be expected to result, as the 8-person teams indicated. Although it was a small sample size, one conjecture could be that the large team size made team members less engaged, and less likely to perform with a concern for excellence (Salas et al., 1999).

The benefit that additional people provide could be related to their contribution to the team's collective memory, as teams have been found to (1) have a larger pool of available information because they can rely on each other's memory, (2) correct the errors of other team members, whereas there is no such process with individuals, and (3) have more effective decision making processes on judgments regarding remembered information (Hinsz, 1990). Within 3-person teams, one factor to achieving this benefit seemed to be balanced communication interactions. As communication has been identified as a key component of team cognition by way of the process of information processing and sharing (Salas et al., 2008), engineering a team structure via resource allocation that facilitates balanced communications could result in improved teamwork and overall performance improvement. The roles that were observed to potentially emerge as a function of resource allocation and their interaction with information were used to guide potential team structures to quantify these findings. These are described in the next chapter, the Phase 2 experiments.

4

Phase 2

The situation of the “Red/Blue” exercise is one in which teams have high uncertainty about the situation, particularly at the beginning of the game. Therefore, using whatever resource or method possible, throughout the game, people will seek to reduce their uncertainty. In this setting, people will act as active information seekers (Gibson, 1972) if we accept the definition of information (according to Information Theory) as “reduction of uncertainty” (Shannon & Weaver, 1949). In relation to SDT, high uncertainty will result in low awareness of the distinction between noise and signal, resulting in low sensitivity. With the addition of information, and the corresponding reduction in uncertainty, the disparity between the signal and noise curves will increase, thereby increasing sensitivity.

This theoretical premise would indicate the ability to effectively acquire and process information as the key to improving a team's sensitivity from its initial uncertain state. In the "Red/Blue" game, the primary resource is the computer, and as a resource, the computer is a source of data allowing a person to acquire and generate information. Based on this role of the computer, for the Phase 2 experiments, resource allocation was defined as the number of computers provided to each team.

Phase 2 focused on the teamwork component of the decision process, in how a team works together to reduce its collective uncertainty. This teamwork was reflected in Δ_{dec} and Δ_{tot}, and the investigation was regarding the factors that influenced how teams could improve upon their initial taskwork. To that extent, where appropriate, the results were normalized to d'_{det}. This was not to diminish the effect of taskwork performance in its effect on overall performance, but the normalization was to emphasize the internal dynamics of the teamwork effect. By normalizing each team to its own d'_{det}, the individual team differences in each team's task performance were able to be subtracted out. The focus thereby was on decision stage performance by visualizing an isolated look at the effect of teamwork on performance, and the factors that most influence this teamwork.

Team size and resource allocation (number of computers per team) were the two variables in Phase 1. Phase 2 experiments extended the results of Phase 1, and were designed to further quantify the effect of resources on teamwork and team performance by holding team size constant, and varying resource allocation (number of computers per team). "Resources" in Phase 2 were computers, and the three conditions for Phase 2 are described as "1-computer," "2-computer," and "3-computer" teams, the labels corresponding to the number of computers that were given to teams in that condition. Varying resource allocation was expected to affect how teams self-organized, their communication interactions, teamwork, and corresponding effect on performance.

4.1 Methodology

4.1.1 Participants

Participants' ages ranged between 18 and 25 years (M = 19.7, mode = 19, median = 19). Table 4.1 provides a summary. 30 participants (18 men and 12 women) participated in 3-computer teams, for a total of 10 teams. 30 participants (24 men and 6 women) participated in 1-computer teams, for a total of 10 teams. 1 team from the 3-computer condition was dropped from the analysis due to a training abnormality. Power analysis confirmed that the sample size was still sufficient. 2-computer team results from Phase 1 were leveraged for Phase 2. Data for the Phase 2 experiments were collected at the Human Factors and Applied Cognitive Engineering Lab at Tufts University. Participants were recruited from the student subject pool at Tufts University. Each participant received $20 per hour for their participation. Table 4.1 provides a summary of the Phase 2 experiments with demographic information.

Table 4.1 *Phase 2 summary*

Number of Computers per Team	Number of Teams	Age	Computer experience (1-5)	Video game frequency (1-5)	Intelligence analysis experience (years)
1	10	19.3	4.1	2.7	0.0
2	15	26.0	4.8	2.6	0.6
3	10	20.3	4.4	2.4	0.0

As in Chapter 3, the provided values in Table 4.1 are averages for the teams in the particular resource allocation. Participants rated their computer experience on a scale of 1-5, with a value of "1" representing "never use a computer" to "5" representing "extremely comfortable with a computer." They also rated how often they play video games on a scale of 1-5, with a value of "1" representing "never," "3" representing "once a week," and "5" representing "daily." And each participant reported how many years of experience they had in intelligence analysis. All participants were comfortable using a computer, and had similar frequency of playing video games. All of the participants in the 1-computer and 3-computer conditions were Tufts University students, and did not have any intelligence analysis experience.

4.1.2 Procedure

All teams followed the same procedure for the game as described in Chapter 2 (General Method).

4.1.3 Variables

The only manipulation (independent variable) was resource allocation, the number of computers that teams were given. Ten teams were given 1 computer, and ten teams were given 3 computers. Teams were randomly selected to belong to either group.

Through these groupings, and leveraging the Phase 1 experiments for the 3-person teams, three team concepts were proposed. In one concept, the team was given only 1 computer, which could potentially promote an organization of two facilitators and one analyst. The second concept is the setup that had been run in Phase 1, in which they are given 2 computers, which promoted the organization of a traditional hierarchy, with one facilitator and two analysts. The third concept attempted to mimic a heterarchical or "flat" organization, in which all three members are each given a computer. This resource allocation did not necessarily promote a facilitator role.

The dependent variables were performance, teamwork, communication interactions, subjective workload, and subjective SA. Performance was measured by the number of hits and false alarms at each stage, in the same way these metrics were collected for Phase 1. As in Phase 1, these metrics were used to calculate *P(H)* and *P(FA)*, and *d'* of SDT, in both the detection phase and decision phase of the exercise, and teamwork as characterized by Δ_{dec} and Δ_{tot}. As described in Chapter 3, Δ_{dec} is defined as the difference between a team's *d'* at the decision stage (d'_{dec}) and its *d'* at the detection stage (d'_{det}).

$$\Delta_{dec} = d'_{dec} - d'_{det}$$

Δ_{tot} is defined as the difference between a team's total (overall) *d'* (d'_{tot}) and its *d'* at the detection stage (d'_{det}).

$$\Delta_{tot} = d'_{tot} - d'_{det}$$

Communication interactions were counted real-time, and verified postgame via video analysis. Self-assessed subjective questionnaires were also collected from each team member. NASA Task Load Index (NASA-TLX) was used to measure individual workload (Hart & Staveland, 1988). Situational Awareness Rating Technique (SART) was used to measure subjective situational awareness (Taylor, 1990). In addition, demographics and background information were posed in the questionnaire. The subjective questionnaires from the Phase 1 experiments yielded some findings, as there were indications from SART ratings that corresponded with performance. NASA-TLX did not yield any significant differences in the Phase 1 experiments, but were used to evaluate whether workload continued to remain relatively constant across the different team concepts. These additional data points were not the primary focus of the research, but were collected as potential supportive evidence to the outcomes.

Table 4.2 summarizes the dependent variables for the Phase 2 experiments, with one independent variable (resource allocation).

Table 4.2 *Independent Variable (IV) and Dependent Variables (DV)*

	IV: Resource Allocation		
	1 Computer (1C)	**2 Computers (2C)**	**3 Computers (3C)**
	10 teams	Completed	10 teams
DV	Number of Hits, False Alarms (detection, decision stages)		
	P(H), P(FA) (detection, decision stages)		
	d'_{det}, d'_{dec}, d'_{tot}		
	Δ_{dec}, Δ_{tot}		
	Communication Interaction Ratio (CIR)		
	NASA-TLX, SART		

4.1.4 Research Questions

The only manipulation for the Phase 2 experiments was resource allocation, which for this phase, was defined as the number of computers provided to the team. The prediction was that the number of computers would promote certain team organizations, teamwork, and corresponding improvements in performance. All other resources, such as projector, map, whiteboard, paper, and stickies, were provided identically for each team concept.

Based on the independent variable (number of computers), three team organization structures were proposed, as illustrated in Figure 4.1.

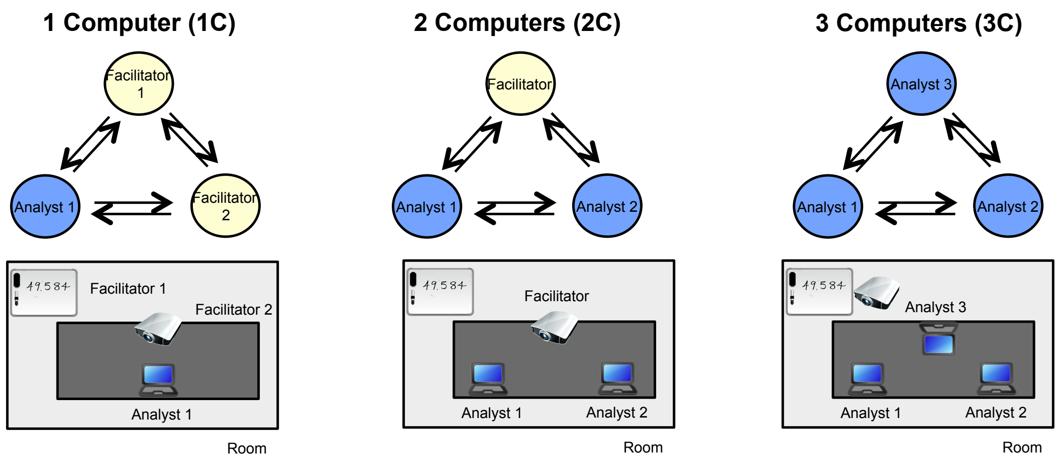


Figure 4.1 *Three team organization structures via resource allocation*

In the "1C" condition, the team was given only 1 computer, and potentially promoted an organization of having two facilitators and one analyst. The "2C" condition was the setup that had been run in Phase 1, in which teams were given two computers, which promoted the organization of one facilitator and two analysts. The "3C" condition attempted to mimic a "flat" organization, in which all three members were equally given computers, and promoted no obvious facilitator role.

Communication interactions amongst team members were expected to vary according to resource allocation, and as teamwork had been found as a byproduct of balanced communication within the team, it was expected that the team organization concept which promoted balanced communication would result in higher levels of teamwork, as defined by

Δ_{dec}, and correspondingly, higher levels of overall improvement in sensitivity, as defined by Δ_{tot}. Overall team performance was also measured within the framework of SDT, with d'_{tot} representing the team's overall ability to correctly discriminate true targets from false targets. Within this framework, various questions were explored.

Q1: What is the effect of resource allocation on team organization, teamwork and performance?

Q2: Which resource allocation will promote highest teamwork and improved performance?

Q3: What is the effect of resource allocation on communication interactions?

Q4: Does the relationship between balanced interactions and teamwork hold amongst different team concepts?

Q5: Does the relationship between improved teamwork (Δ_{dec}) and improved performance (Δ_{tot}) hold across resource allocation?

Phase 2 aimed to evaluate which team concept would give rise to better teamwork and performance improvement, as defined by Δ_{dec} and Δ_{tot}. The hypothesis was that the 3-computer teams would have the highest Δ_{dec} and Δ_{tot} and have the most balanced communication interactions. However, it was also recognized that there could be a variety of potential outcomes. Table 4.2 summarizes nine potential outcomes of Phase 2 with respect to these metrics. As mentioned in the introduction, there are many thoughts and theories about teamwork and team performance, and each concept would be tested by the various potential outcomes, as noted in the table. Performance outcomes were evaluated according to the above descriptions, as defined by Δ_{dec} and Δ_{tot}.

Table 4.3 *Potential outcomes of Phase 2*

	1C		2C		3C	Outcome
1	F1, A1, F2	<	F1, A1, A2	<	A3, A1, A2	3C mimics heterarchical teams, which have been found to result in better team performance (Bowers, Urban, and Morgan, 1992) and more accuracy than hierarchical teams, which could be due to better exchange of knowledge (Schraagen, Veld, & De Koning, 2010).
2	F1, A1, F2	>	F1, A1, A2	>	A3, A1, A2	In 1C, all team members are viewing the same information at all times (shared display), which can improve shared mental models and team SA (Bolstad and Endsley, 2000), and increased shared information may give rise to increased communication (Mesmer-Magnus and DeChurch, 2009)
3	F1, A1, F2	<	F1, A1, A2	>	A3, A1, A2	2C mimics traditional hierarchical teams, and in stable and predictable environments (which can be the case when all data is analyzed forensically – no new data introduced during the game), hierarchy is recommended (Schraagen and Rasker, 2003).
4	F1, A1, F2	>	F1, A1, A2	<	A3, A1, A2	Traditional hierarchical teams are ones in which team members pass information to the "leader" but not to each other, resulting in poor information sharing (Schraagen et al., 2010). Teams that share task- and team-related knowledge better can have better performance (Cooke, Salas, Kiekel, and Bell, 2004).
5	F1, A1, F2	=	F1, A1, A2	<	A3, A1, A2	Heterarchical teams freely exchange information with one another to reach a common goal (Humphrey, Hollenbeck, Meyer, and Ilgen, 2002), and have been shown to be faster, more accurate, and share more information (Lewelling and Nissen, 2007; Martin and McEver, 2008).
6	F1, A1, F2	=	F1, A1, A2	>	A3, A1, A2	Hierarchical teams can communicate more than nonhierarchical teams when there is task specialization, when no team member has an overview of the whole task (Urban, Bowers, Monday, and Morgan, 1995).
7	F1, A1, F2	>	F1, A1, A2	=	A3, A1, A2	In a team setting, regardless of resource allocation, one person will tend to emerge in a facilitator role, resulting in 2C and 3C to effectively be the same. Team leadership (shared facilitation) can help team performance in "improvisational" organizational units (Klein, Ziegert, Knight, and Xiao, 2006)
8	F1, A1, F2	<	F1, A1, A2	=	A3, A1, A2	In 1C, there could be too little task specialization due to two people not having computers, which results in poorer performance (Urban, et al., 1992), and in both 2C and 3C, one person emerges in a facilitator role.
9	F1, A1, F2	=	F1, A1, A2	=	A3, A1, A2	There is no impact of resource allocation. When there is a balance between task specialization and load balancing, there is no difference in performance (Urban, Weaver, Bowers, and Rhodenizer, 1996).

The first potential outcome in Table 4.3 describes a result in which the 3-computer (3C) condition outperforms the 2-computer (2C) and 1-computer (1C) conditions. This potential outcome could confirm results that found heterarchical teams, which the 3-computer (3C) condition mimics, resulted in better team performance (Bowers, Urban, & Morgan, 1992) and more accuracy than hierarchical teams, which is mimicked by the 2-computer (2C) condition. The 1-computer (1C) condition mimicked a version of a hierarchical team. In addition, allowing each team member their own source of information could result in better exchange of knowledge (Schraagen et al., 2010), as knowledge-redundant teams (those with more shared knowledge) share more information than less knowledge-redundant teams (Mesmer-Magmus & DeChurch, 2009).

The second potential outcome describes a result in which the 1-computer (1C) condition outperforms the 2-computer (2C) and 3-computer (3C) conditions. As there is only one computer, it can become the shared display for the team, and may improve shared mental models and team SA (Bolstad & Endsley, 2000). And just as the 3-computer (3C) could promote knowledge-redundancy, having one computer could just as likely promote knowledge-redundancy (Mesmer-Magnus & DeChurch, 2009).

The third potential outcome describes a result in which the 2-computer (2C) condition outperforms the 1-computer (1C) and 3-computer (3C) conditions. Schraagen and Rasker (2003) found that a hierarchy is recommended when the situation is stable and predictable. The "Red/Blue" game, as a forensic data analysis exercise, could be one for which that description applies. As such, and as the 2-computer (2C) condition most resembles the traditional hierarchical structure, if those assumptions are met, the 2-computer (2C) condition could be the condition that outperforms the others.

The fourth potential outcome is one in which the 2-computer (2C) condition is the least performing condition. This could be the outcome if the traditional hierarchical structure holds true as far as communication patterns, where team members pass information to the "leader" but not to each other, resulting in poor information sharing (Schraagen et al., 2010). Teams that share task and team related knowledge have been found to have better performance (Cooke et al., 2004).

The fifth potential outcome is one in which the 3-computer (3C) condition outperforms the other two conditions, but there is no distinction between the 1-computer (1C) and 2-computer (2C) conditions, which are forms of a potential hierarchical structure. As the 3-computer (3C) condition potentially would best allow free exchange of information (due to lack of hierarchy), it could allow teams to share more information and be more accurate (Hollenbeck et al. 2002; Lewelling & Nissen, 2007; Martin & McEver, 2008).

The sixth potential outcome is the opposite of the fifth potential outcome, in which the 1-computer (1C) and 2-computer (2C) conditions equally outperform the 3-computer (3C) condition. With this potential outcome, the benefit of a hierarchical structure could be demonstrated, and be consistent with results that showed hierarchical teams can communicate more than nonhierarchical teams when no team member has an overview of the whole task (Urban et al., 1995).

The seventh potential outcome is one in which the 1-computer (1C) condition outperforms the other two. Such a result would show that resource allocation is less of a factor than the emergence of a facilitator. In a team setting, regardless of resource allocation, in this potential outcome, one person would tend to emerge in a facilitator role, resulting in the 2-computer (2C) and 3-computer (3C) conditions to effectively be the same. The 1-computer (1C) condition would then mimic a “shared facilitation” structure, or team leadership, which can help team performance in "improvisational" organizational units (Klein, Ziegert, Knight, & Xiao, 2006).

The eighth potential outcome is one in which there is no distinction between the 3-computer (3C) and 2-computer (2C) conditions, but they both outperform the 1-computer (1C) condition. As in the seventh potential outcome, this potential outcome would have one person emerge in a facilitator role regardless of the resources, resulting in the 2-computer (2C) and 3-computer (3C) conditions to effectively be the same. And with the 1-computer (1C) condition, there could be very little task specialization due to two people not having computers, which would result in poorer performance (Urban et al., 1992).

Lastly, the ninth potential outcome is one in which there is no difference amongst the three conditions. Such a result could indicate that there is no impact of resource allocation, and that people are able to successfully adapt to whatever situation in which they are placed. This result could indicate that the number of people on a team, not the number of resources, is a factor that drives performance.

4.2 Results

Table 4.4 summarizes the results for the three different conditions, using the two-stage decision process model presented in Chapter 2, Section 2.3.3. Full descriptives for all of the dependent variables are shown in Appendices E-G. Performance, as described by SDT, was measured in terms of hits and false alarms, which were used to calculate the probability of hits and probability of false alarms at each stage, as well as at the overall (or total) level. These values were used to calculate *d'* at each of the stages.

Table 4.4 *Phase 2 results summary (mean values)*

	# of Computers	Place marks	Actions	Hits	FAs	P(H)	P(FA)	d'
Detection	1	14.2		5.4	8.8	0.54	0.05	**1.85**
	2	14.5		5.0	9.5	0.50	0.05	**1.68**
	3	19.4		5.1	14.3	0.51	0.08	**1.52**
Decision	1		7.0	3.9	3.1	0.71	0.37	**0.99**
	2		5.5	3.5	2.0	0.67	0.22	**1.37**
	3		6.6	3.8	2.8	0.73	0.23	**1.49**
Total	1			3.9	3.1	0.39	0.02	**1.93**
	2			3.5	2.0	0.35	0.01	**1.97**
	3			3.8	2.8	0.38	0.02	**1.92**

Figure 4.2 presents these results as a traditional Receiver Operator Characteristic (ROC) plot, which is a graphical representation of the point

estimate of the sensitivity of the various teams at the detection and decision stages, and at the overall level. The values for the probability of hits ($P(H)$) and false alarms ($P(FA)$) from Table 4.4 are plotted.

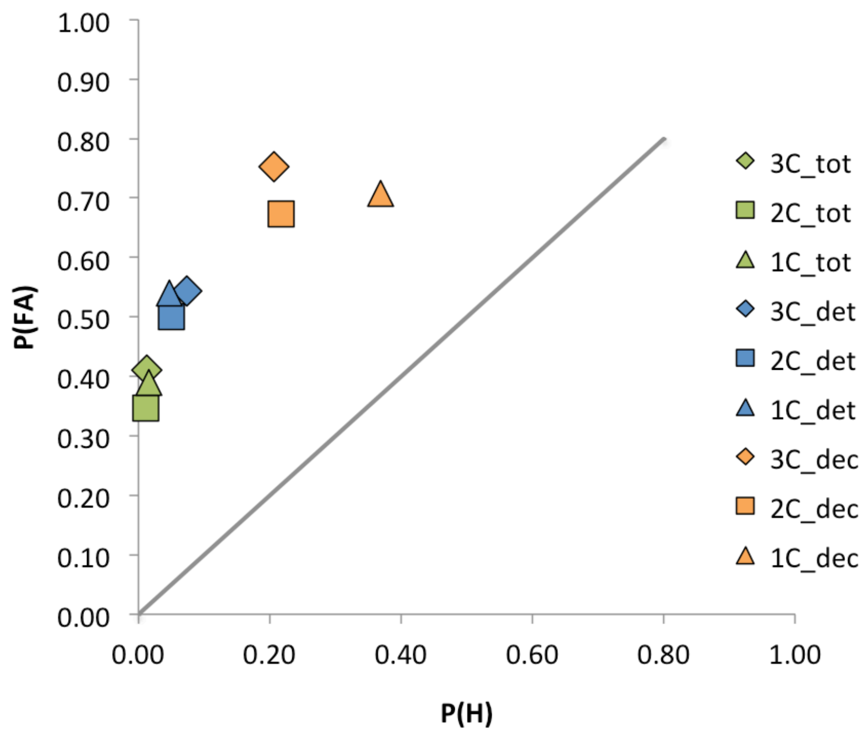


Figure 4.2 *Phase 2 results on ROC graph*

Regarding notation for the figures in Chapter 4, "1C" refers to the 1-computer teams, "2C" refers to the computer teams, and "3C" refers to the 3-computer teams. The labels next to each of the team designations refer to the stage: "det" is detection stage, "dec" is decision stage, and "tot" is the total or overall result.

The number of placemarks, actions, hits, and false alarms were measured in the same manner as they were in Phase 1, described in Chapter 3, section 3.2. Likewise, *P(H)*, *P(FA)*, and *d'* for each of the stages were calculated in the same way as they were in Phase 1.

Phase 2 focused on the improvement of sensitivity, leveraging the findings from Phase 1, by examining both Δ_{dec} and Δ_{tot}. In the results from Phase 1, a high Δ_{dec} showed that the aggregate had improved knowledge in some way, while many underachieving teams had results that showed that they could be no different than individuals, perhaps reflecting a lack of teamwork. And while high Δ_{dec} showed improvement of knowledge, as Δ_{dec} values became more negative, it seemed to reflect negative teamwork in counter productive relationships amongst team members.

In addition, higher levels of Δ_{dec} increased the likelihood that a team would have improved overall performance relative to its performance at the detection stage, shown by the positive correlation between Δ_{dec} and Δ_{tot} shown in Figure 3.13. Phase 1 results showed that regardless of a team's taskwork, as captured by d'_{det}, improved teamwork (high Δ_{dec}) could result in improvement in performance upon its taskwork performance (high Δ_{tot}). Phase 2 data analyses focused on these aspects of the process.

As Δ_{dec} and Δ_{tot} were the primary metrics of interest, in appropriate cases, the data were normalized to better visualize these effects. Each team's data were normalized to the team's performance at the detection stage, d'_{det}. Again, it is acknowledged that there can be a variety of individual factors that affect d'_{det}, which is reflected in the variety of ways that teams approached their task. And as it is understood that the detection stage performance reflects the taskwork component, normalizing to that value simply allowed a better visualization of the teamwork components, Δ_{dec} and Δ_{tot} by subtracting out individual team differences in how each team carried out their taskwork.

As in Phase 1, analyses of the data were similarly performed through one-way independent measures ANOVA. When the assumption of normality or homogeneity of variances was not met, the nonparametric test, Kruskal-Wallis one-way analysis of variance was performed. The Shapiro-Wilk test was used to test for normality, and Levene's test was used to test for homogeneity of variances. When statistically significant results were discovered, the Tukey HSD or Dunn test was used to perform post-hoc paired comparisons.

4.2.1 One-Computer Teams

Figure 4.3 shows the results of all the teams from the 1-Computer (1C) condition.

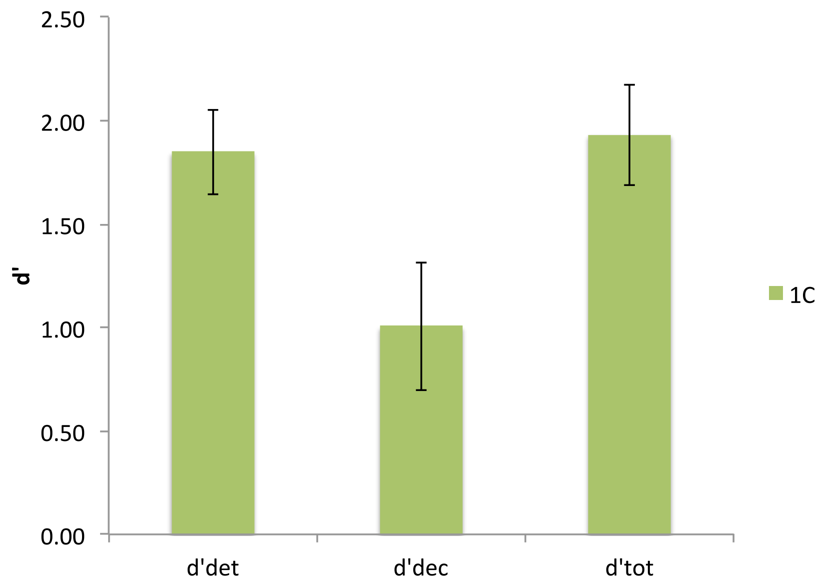


Figure 4.3 *1-Computer (1C) team results (d' values, mean)*

The mean d'_{det} was 1.85 (SD = .50), the mean d'_{dec} was 1.01 (SD = .80), and the mean d'_{tot} was 1.93 (SD = .54). The mean Δ_{dec} was -0.84 (SD = .94) and the mean Δ_{tot} was 0.08 (SD = .40).

Figure 4.4 shows the results for all of the 1-computer teams as a line plot, which better illustrates each team's performance at the various stages. As stated earlier, of primary interest is in understanding how a team can arrive at an overall gain in performance from its taskwork at the detection stage, versus teams that have loss in performance from its taskwork. As such, Figure 4.4 shows the teams with $\Delta_{tot} \geq 0$ as blue lines, and the teams with $\Delta_{tot} < 0$ as orange lines. In addition, each team is identified by its numerical identification.

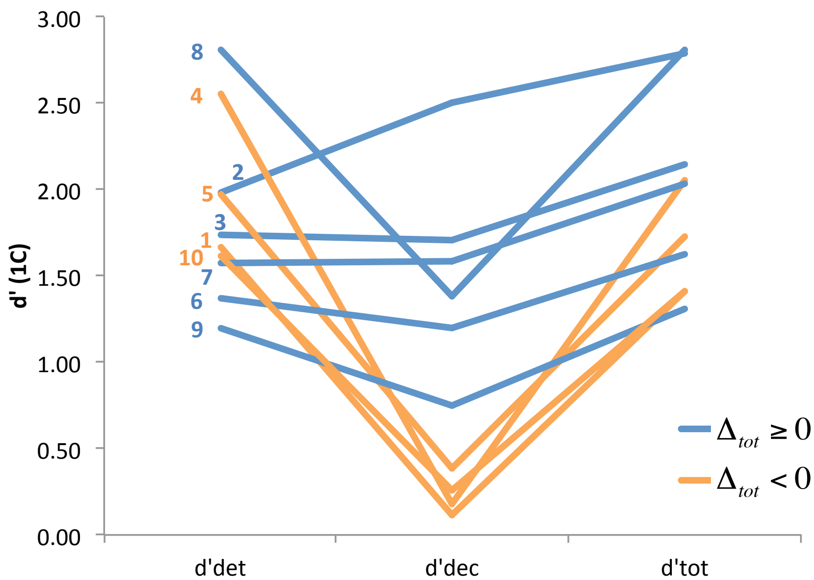


Figure 4.4 *1-Computer (1C) team results (d' values) identified by Δ_{tot}*

For teams in the 1-computer condition, there were six teams (team numbers 2, 3, 6, 7, 8, and 9) with Δ_{tot} greater than or equal to zero (60%), and four teams (team numbers 1, 4, 5, and 10) with a negative Δ_{tot} (40%). As Figure 4.4 shows, the four teams with negative Δ_{tot} (team numbers 1, 4, 5, and 10) had the lowest d'_{dec} values.

To better visualize Δ_{dec} and Δ_{tot}, the normalized results (normalized to d'_{det}) are shown in Figure 4.5. The team identification numbers can be compared with the corresponding line plot in Figure 4.4.

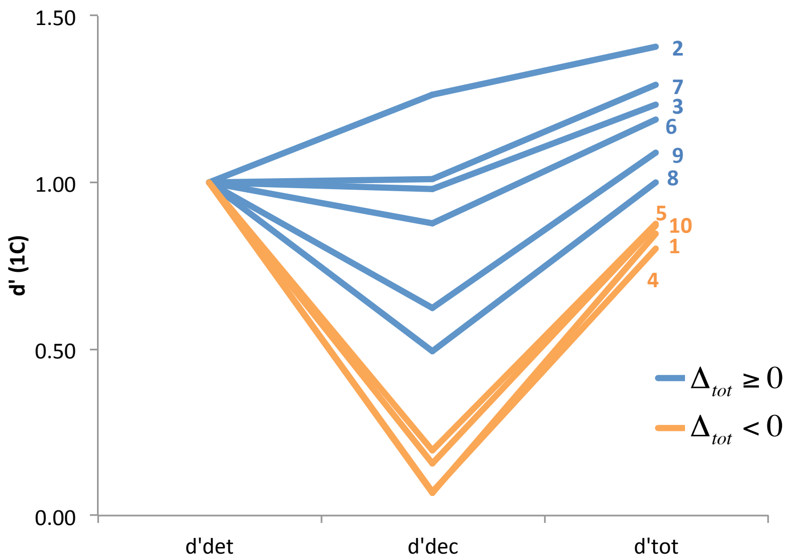


Figure 4.5 *Normalized 1-Computer (1C) team results (d' values) identified by* Δ_{tot}

Figure 4.5 shows results that indicate, for the 1-computer (1C) condition, 60% of the teams were able to use their decision stage to result in an overall improvement, or minimally no loss, in performance from their detection stage performance. The indication is that 40% of the teams were not able to achieve teamwork in the decision stage that allowed them to improve upon their taskwork performance. Rather, their lack of teamwork (as indicated by the low d'_{dec} values in Figure 4.5) may have resulted in a corresponding loss of overall performance compared to their detection stage performance.

4.2.2 Two-Computer Teams

Figure 4.6 shows the results of all the teams from the 2-Computer (2C) condition.

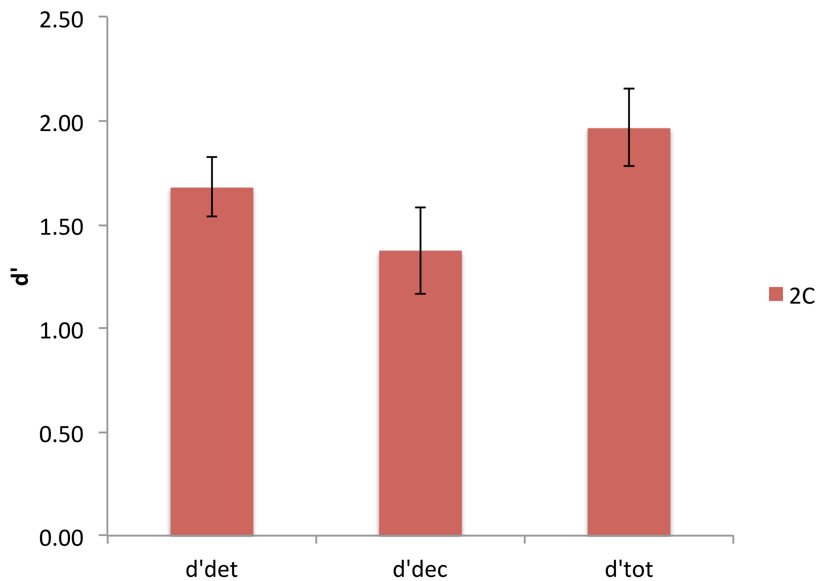


Figure 4.6 *2-Computer (2C) team results (d' values, mean)*

The mean d'_{det} was 1.68 (SD = .45), the mean d'_{dec} was 1.37 (SD = .66), and the mean d'_{tot} was 1.97 (SD = .59). The mean Δ_{dec} was -0.31 (SD = .56) and the mean Δ_{tot} was 0.29 (SD = .32). Figure 4.7 shows the results for all of the teams as a line plot. The teams with $\Delta_{tot} \geq 0$ are shown as blue lines, and the teams with $\Delta_{tot} < 0$ as orange lines. In addition, each team is identified by its numerical identification.

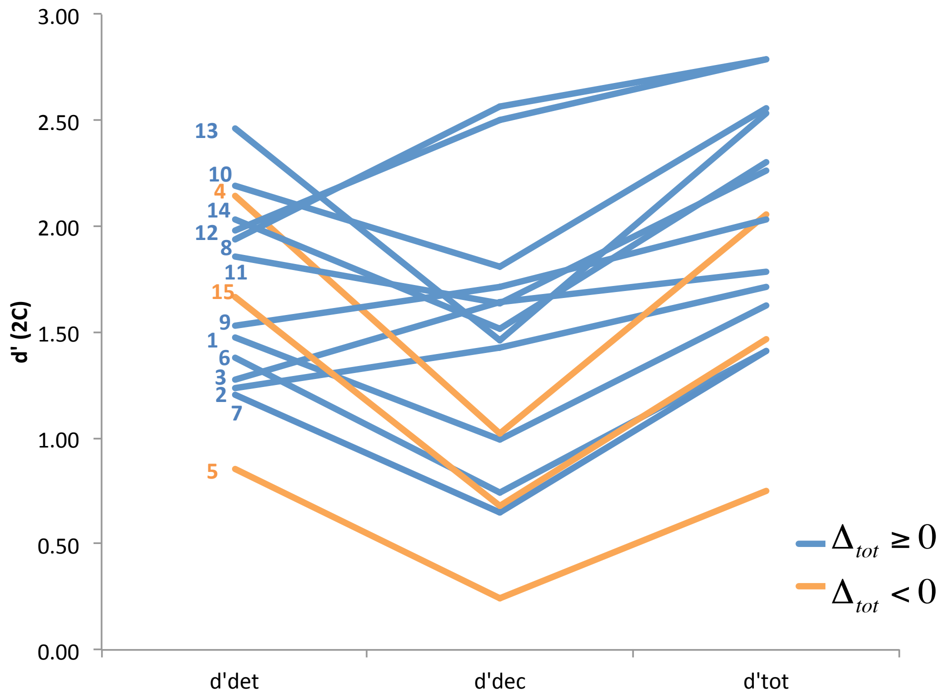


Figure 4.7 *2-Computer (2C) team results (d' values) identified by Δ_{tot}*

As Figure 4.7 shows, for teams in the 2-computer condition, there were twelve teams (team numbers 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, and 14) with Δ_{tot} greater than or equal to zero (80%), and three teams (team numbers 4, 5, and 15) with a negative Δ_{tot} (20%).

To better visualize Δ_{dec} and Δ_{tot}, the normalized results (normalized to d'_{det}) are shown in Figure 4.8. The team identification numbers can be compared with the corresponding line plots in Figure 4.7.

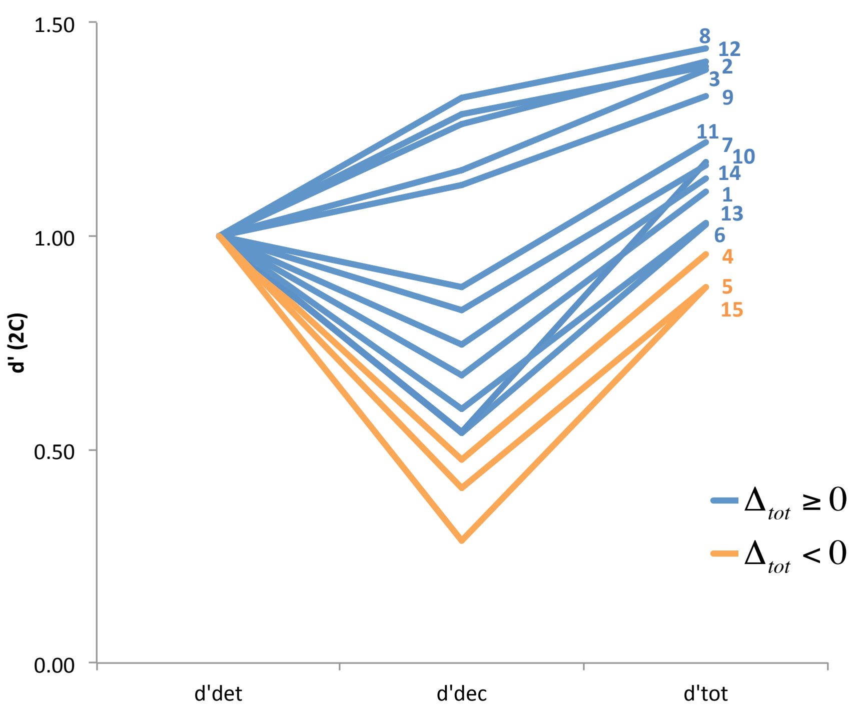


Figure 4.8 *Normalized 2-Computer (2C) team results (d' values) identified by Δ_{tot}*

These results indicate that for the 2-computer (2C) condition, 80% of the teams were able to use their decision stage to result in an overall improvement, or minimally no loss, in performance from their detection stage performance. The indication is that 20% of the teams in this condition were not able to achieve teamwork in the decision stage that allowed them to improve upon their taskwork performance. Rather, their lack of teamwork (as indicated by the low d'_{dec} values in Figure 4.8) may have resulted in a corresponding loss of overall performance compared to their detection stage performance.

4.2.3 Three-Computer Teams

Figure 4.9 shows the results of all the teams from the 3-Computer (3C) condition.

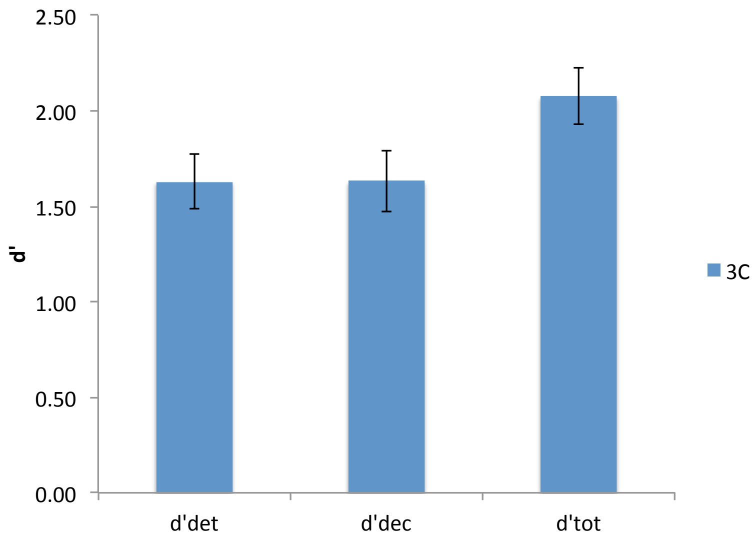


Figure 4.9 *3-Computer (3C) team results (d' values, mean)*

The mean d'_{det} was 1.63 (SD = .45), the mean d'_{dec} was 1.63 (SD = .51), and the mean d'_{tot} was 2.08 (SD = .46). The mean Δ_{dec} was .003 (SD = .51) and the mean Δ_{tot} was 0.45 (SD = .23). Figure 4.10 shows the results for all of the teams as a line plot. In addition, each team is identified by its numerical identification.

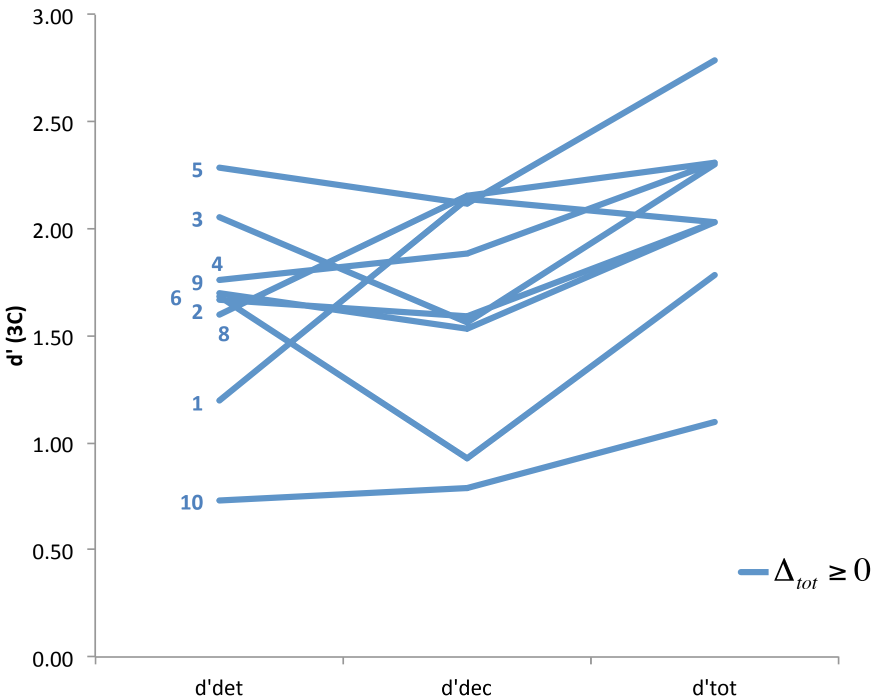


Figure 4.10 *3-Computer (3C) team results (d' values) identified by Δ_{tot}*

As Figure 4.10 shows, for teams in the 3-computer condition, all ten teams achieved Δ_{tot} greater than or equal to zero (100%).

To better visualize Δ_{dec} and Δ_{tot}, the normalized results (normalized to d'_{det}) are shown in Figure 4.11. The team numerical identification can be compared with the corresponding line in Figure 4.10.

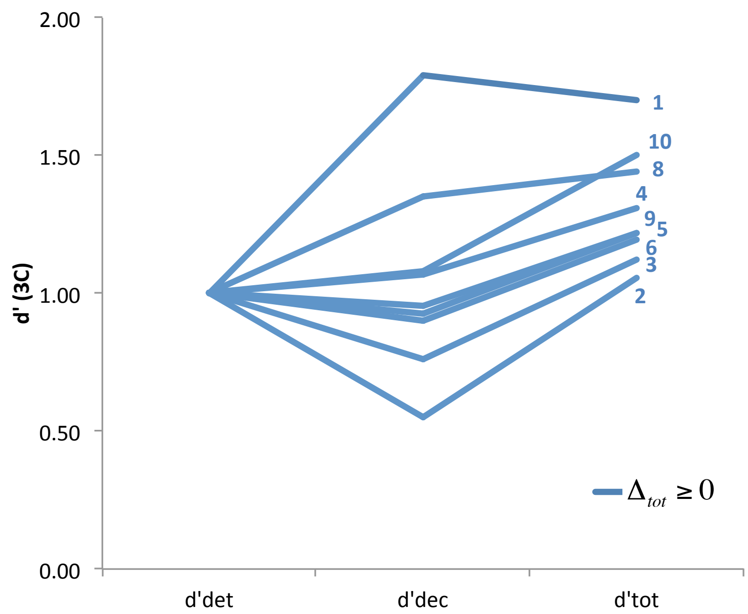


Figure 4.11 *Normalized 3-Computer (3C) team results (d' values) identified by Δ_{tot}*

These results indicate that for the 3-computer (3C) condition, 100% of the teams were able to use their decision stage to result in an overall improvement, or minimally no loss, in performance from their detection stage performance. The indication is that in this condition, all teams were able to achieve some level of teamwork in the decision stage that allowed all of them to improve, or minimally have no loss, upon their taskwork performance.

4.2.4 Impact of Resources

Figure 4.12 summarizes the *d'* values for each condition for each stage of the game.

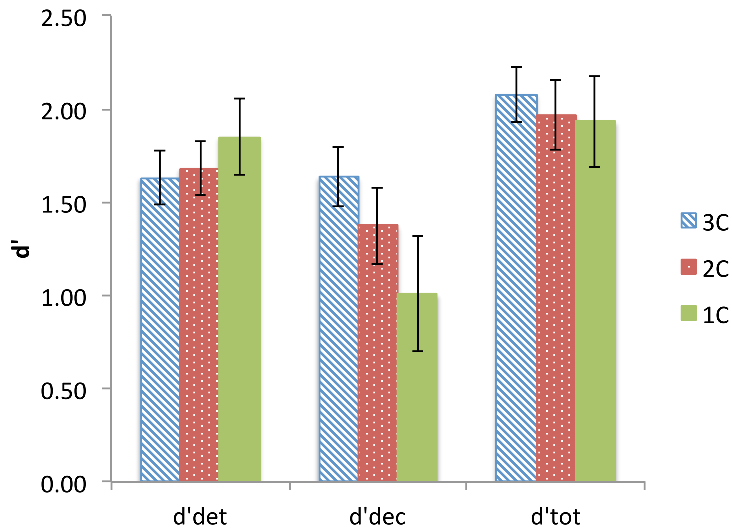


Figure 4.12 *Phase 2 results (d' values) categorized by resource allocation (mean)*

For d'_{det}, d'_{dec}, and d'_{tot}, one-way independent measures ANOVA showed that there were not significant main effects for resource allocation [$F(2, 31) = .60$, $p = .554$; $F(2, 31) = 2.12$, $p = .137$; $F(2, 31) = .18$, $p = .835$]. However, as stated in the introduction of this chapter, the emphasis was not in finding statistical differences in the *d'* values as separate entities. This was because the intent of the resource allocation manipulation was to examine the progression of performance, and how a team's performance at the decision stage allows an improvement in overall performance based on its detection stage performance. As such, Figure 4.13 shows the results as a line graph, in which this becomes more evident.

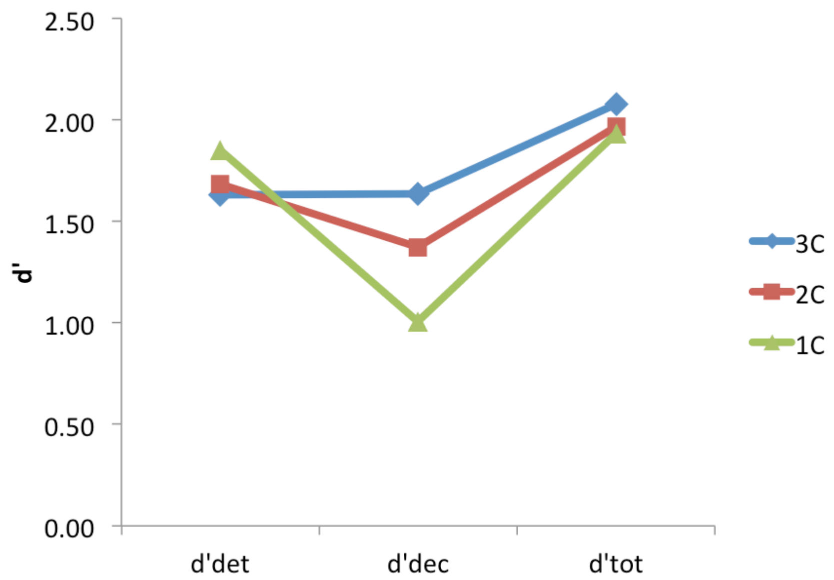


Figure 4.13 *Phase 2 results (d' values) categorized by resource allocation (mean)*

Results showed that the impact of increasing resources resulted in higher d'_{dec}, Δ_{dec}, and Δ_{tot}. Although the difference in d'_{dec} was not statistically significant, the line plot shows a clear indication of 3-computer teams using their detection stage information most effectively in their decision stage performance. To better visualize this observation, a normalized plot of Δ_{dec} and Δ_{tot} by resource allocation is shown in Figure 4.14.

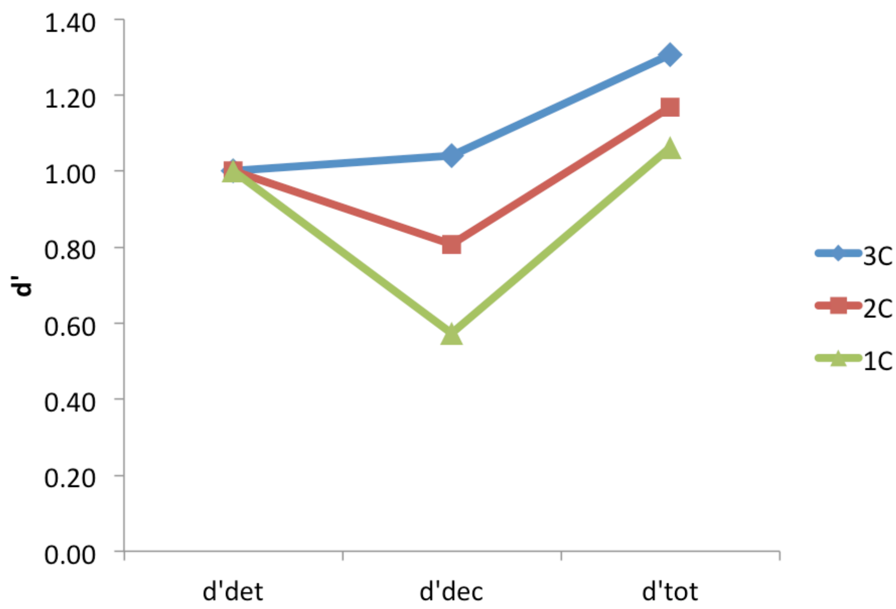


Figure 4.14 *Normalized Phase 2 results (d' values) categorized by resource allocation (mean)*

Figure 4.14 shows a clear indication of 3-computer teams most effectively using their decision stage to improve upon their detection stage to affect their overall performance. In evaluating the normalized results, for d'_{dec}, a one-way independent measures ANOVA showed that there was a statistically significant main effect for resource allocation, $F(2, 31) = 3.63$, $p = .038$. Likewise, for the normalized d'_{tot} results, a one-way independent measures ANOVA showed that there was a statistically significant main effect for resource allocation, $F(2, 31) = 3.55$, $p = .041$. Tukey's HSD tests showed that the normalized d'_{dec} and d'_{tot} values for 3-computer teams were statistically significantly higher than 1-computer teams.

4.2.4.1 Delta decision (Δ_{dec}) and Delta total (Δ_{tot})

To illustrate how the teams of each condition are improving upon their detection stage performance (taskwork), Figure 4.15 shows a plot of Δ_{dec} and Δ_{tot} for each condition.

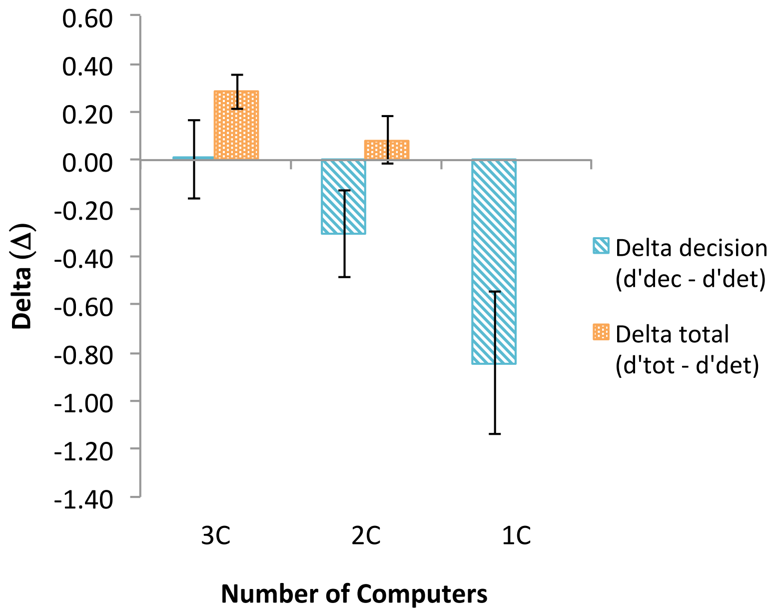


Figure 4.15 *Delta decision (Δ_{dec}) and Delta total (Δ_{tot}) by resource allocation*

For Δ_{dec} (d'_{dec} - d'_{det}), one-way independent measures ANOVA showed that there was a statistically significant main effect for number of computers, $F(2, 31) = 3.81$, $p = .033$. Tukey's HSD tests showed that Δ_{dec} for 3-computer teams was statistically significantly higher than 1-computer teams, showing that in order to have consistently improved performance with 3-person teams, giving each person their own computer is the best configuration. The mean Δ_{dec} for 1-computer teams was -.84 (SD = .94), the mean Δ_{dec} for 2-computer teams was -.31 (SD = .56), and the mean Δ_{dec} for 3-computer teams was .003 (SD = .51).

For Δ_{tot} ($d'_{tot} - d'_{det}$), one-way independent measures ANOVA showed that while not significant, there was a strong indication of main effect for number of computers, $F(2, 31) = 2.96$, $p = .067$. The mean Δ_{tot} for 1-computer teams was .08 (SD = .40), the mean Δ_{tot} for 2-person teams was .29 (SD = .32), and the mean Δ_{tot} for 3-computer teams was .45 (SD = .23).

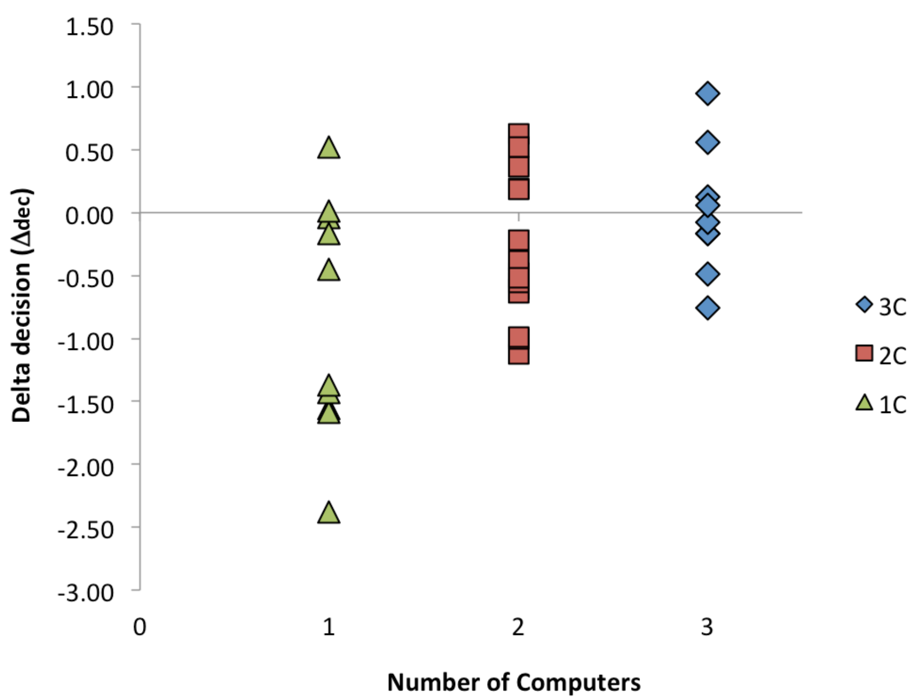


Figure 4.16 *Delta decision (Δ_{dec}) by resource allocation*

As shown in Figure 4.16, there was a significant correlation between the number of computers and Δ_{dec}, $r(34) = .437$, $p = .010$. Likewise, Figure 4.17 shows that there was a significant correlation between the number of computers and Δ_{tot}, $r(34) = .399$, $p = .019$. These results indicate the impact that the number of resources has on a team's ability to improve its decision stage performance from its detection performance, as well as its ability to improve its overall performance from its detection stage performance.

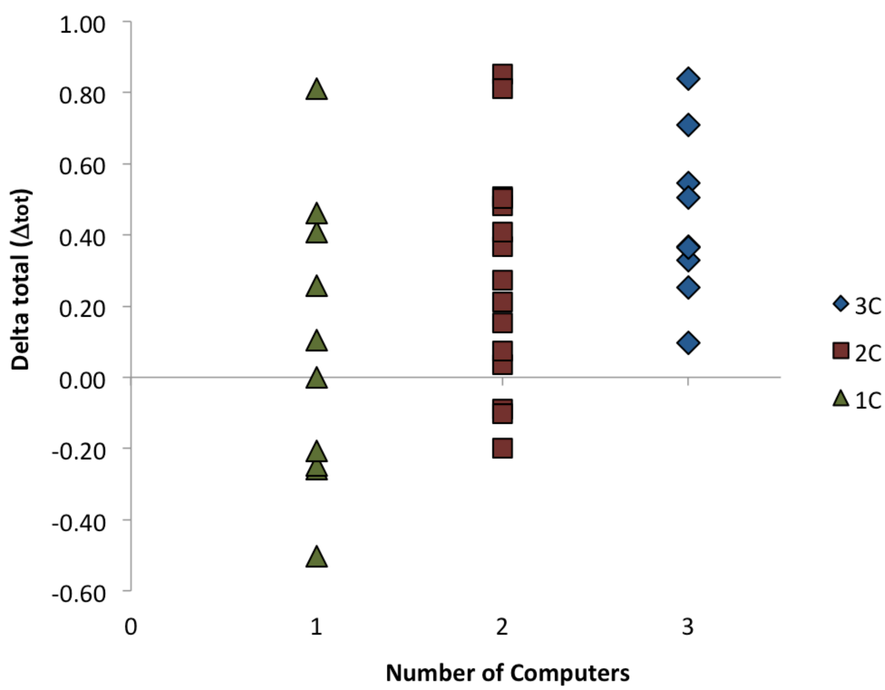


Figure 4.17 *Delta total (Δ_{tot}) by resource allocation*

As in Phase 1, higher levels of Δ_{dec} increased the likelihood that a team would have improved overall performance relative to its performance at the detection stage. The results from Phase 2 continued to show that regardless of a team's taskwork, as captured by d'_{det}, improved teamwork (high Δ_{dec}) can result in improvement in performance upon its taskwork performance (high Δ_{tot}). This is demonstrated by the strong correlation between Δ_{tot} and Δ_{dec} in the Phase 2 results, $r(34) = .955$, $p = .001$, as shown in Figure 4.18.

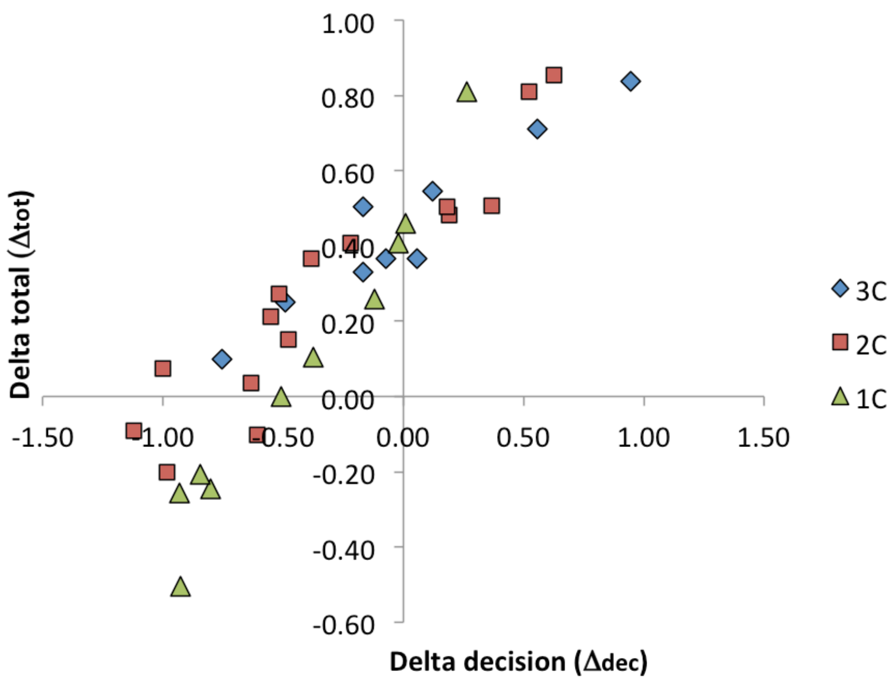


Figure 4.18 *Correlation between Delta total (Δ_{tot}) and Delta decision (Δ_{dec}) in Phase 2*

4.2.5 Communication Interactions

The Communication Interaction Ratio (CIR) was the ratio relating the number of interactions, simple utterances, in a link between particular team members. This ratio has been described earlier in Chapter 3, section 3.2.4. A ratio close to the value of "1" was indicative of balanced interactions, and values greater than "1" was indicative of unbalanced interactions.

As expected, communication interactions in Phase 2 had significant correlation with resource allocation, $r(27) = -.441$, $p = .021$. And as resource allocation had been shown to have positive correlation with Δ_{dec}, Figure 4.19 shows the corresponding relationship between CIR and Δ_{dec} as a function of resource allocation.

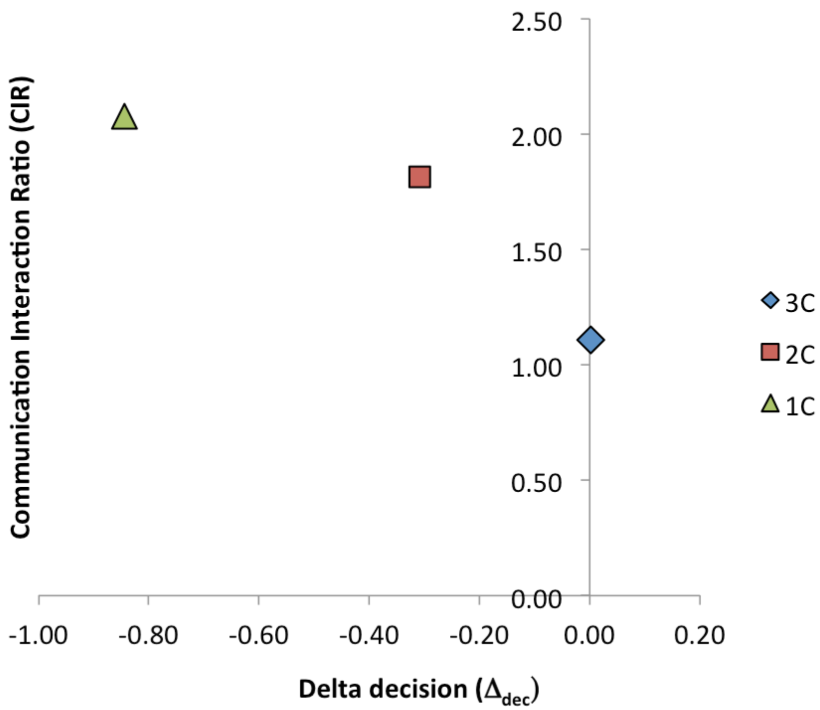


Figure 4.19 *Communication Interaction Ratio (CIR) by Delta decision (Δ_{dec}) (mean)*

Although CIR had significant correlation with resource allocation, and resource allocation had significant correlation with Δ_{dec}, there was not a significant correlation of CIR to Δ_{dec}, $r(27) = -.061$, $p = .732$, due to the large variances in CIR with the 1-computer teams. The mean CIR for 1-computer teams was 2.08 (SD=1.32), the mean CIR for 2-computer teams was 1.69 (SD=.60), and the mean CIR for 3-computer teams was 1.11 (SD=.12).

As the standard deviation values for each condition indicate, rather than simply viewing the means, examining the reduction in variance as a function of resource allocation shows the clearer indication of the impact of resource allocation on Δ_{dec}. The reduction in variance as the number of computers increased is clearly illustrated in Figure 4.20.

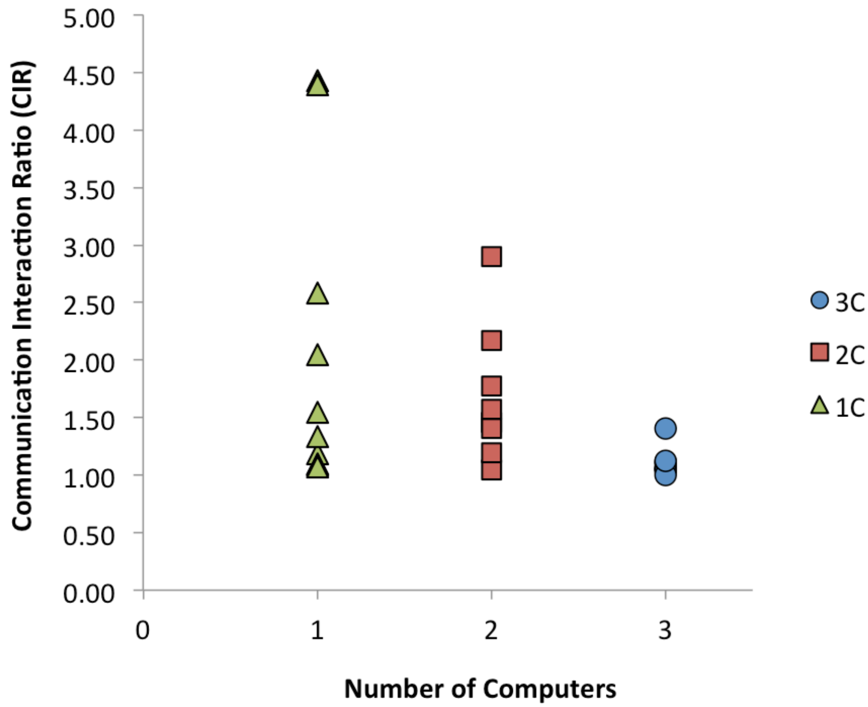


Figure 4.20 *Communication Interaction Ratio (CIR) by resource allocation*

Figure 4.19 shows the trend, and Figure 4.20 shows the impact. As would be expected, the 1-computer condition had the greatest variance, and the 3-computer condition had the least variance. In the 1-computer condition, the communication interactions were relatively unpredictable, shown by the large variance (1.75). CIR became more predictable in the 2-computer condition, with a smaller variance (.36). And the CIR became most predictable in the 3-computer condition, shown by the very small variance (.01). This demonstrates the powerful effect of resource allocation in establishing predictable communication behaviors.

4.2.6 NASA-TLX and SART

The Kruskal-Wallis test showed that workload (NASA-TLX) had statistically significant differences across resource allocation, $\chi^2(2, N=34) = 6.73$, $p = .035$. The mean workload for 1-computer teams was 3.97 (SD = .26), the mean workload for 2-computer teams was 4.33 (SD = .26), and the mean workload for 3-computer teams was 4.29 (SD = .56). Post-hoc pairwise comparisons showed that the 2-computer condition mean workload was statistically significantly higher than the 1-computer condition, but there was not a significant difference between the 1-computer and 3-computer conditions, nor was there between the 2-computer and 3-computer conditions.

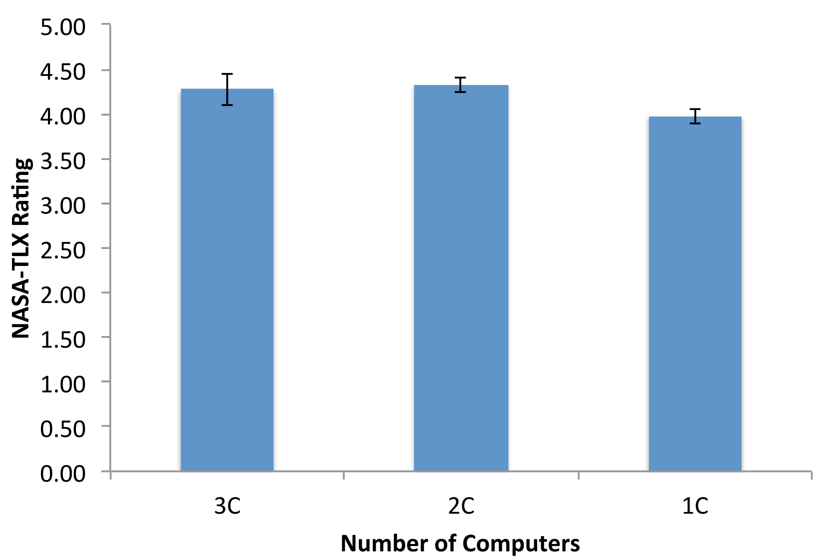


Figure 4.21 *NASA-TLX ratings by resource allocation (mean values)*

Overall, the amount of experienced workload did not vary extremely across resource allocation, as the mean values indicate. The results do show that the addition of one or two more computers had a significant effect, but not between two to three computers.

For the 1-computer teams, there were two people without computers, and the workload results are consistent with anecdotal evidence. At the beginning of the game, in some teams, the two people without computers were observed to be fairly active in writing down information, operating the map or whiteboard, but as the game progressed, they eventually all gravitated towards the one computer. Instead of working in a facilitator role using the whiteboard, map, or projector, the two people without computers spent the majority of their time seated next to the computer. Although the workload results from Phase 1 showed no statistically significant differences, the results from Phase 2 showed a difference. The indication is that when the number of people without computers outnumber those with computers, significant differences in workload can occur. This is intuitively explainable, in that when the number of people without computers outnumber those with computers, there is not enough information being generated to keep the non-computer people busy.

Further examining the variances in the workload results reveals the effect of the computer on homogeneity of experience. A plot of all the workload ratings for every member of each condition is shown in Figure 4.22.

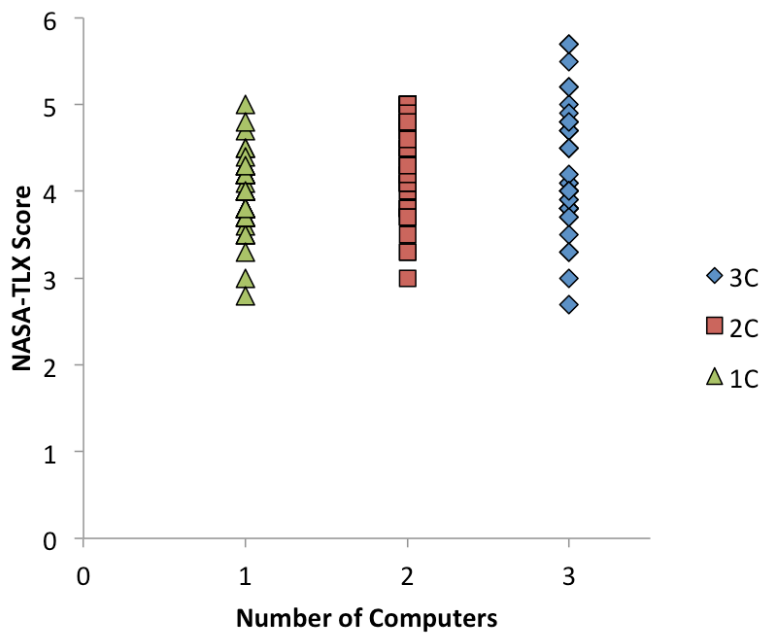


Figure 4.22 *NASA-TLX ratings by resource allocation (all teams)*

Figure 4.22 shows the reduction in variance as the number of computers decreases. The average variance for the 1-computer condition was 0.21, for the 2-computer condition was 0.26, and for the 3-computer condition was 0.50. This reveals that potentially, the effect of the one computer as the lone source of information drove the 1-computer participants to all have the same experience. The computer is what enjoined all the participants into a uniform experience, whereas in the 3-computer condition, when each participant had their own computer, they each interacted with the information in their own unique ways, which resulted in the greater variance. However, this uniqueness of experience with the information promoted by the 3-computer condition was what resulted in the highest levels of teamwork and performance, while the 1-computer condition, which promoted the most similar experiences, resulted in the lowest reported workload.

As a matter of further comparison, Figure 4.23 shows the NASA-TLX ratings for all three conditions along with the individual participant (1p) workload ratings from Phase 1. As with many of the comparisons with the individual participant condition, it could depict the baseline workload characteristics of an individual operating one computer.

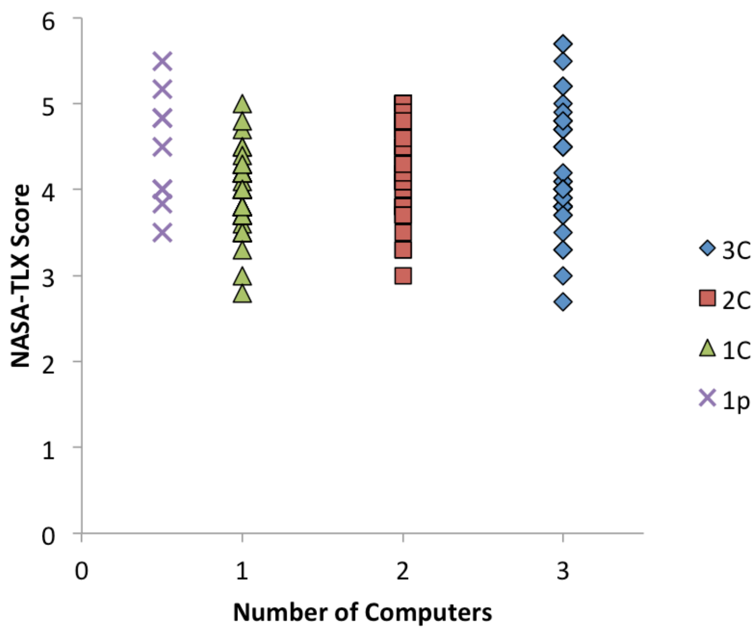


Figure 4.23 *NASA-TLX ratings by resource allocation, with individual participant results from Phase 1 (all teams)*

Figure 4.23 provides added indication of people's tendency of having their own unique way of interacting with the information, given their own computer. It resulted in a higher average amount of workload, and a larger variance (.46) than other conditions.

The Kruskal-Wallis test showed that subjective situational awareness (SART) had indications of statistically significant differences across resource allocation, $\chi^2(2, N=34) = 5.09$, $p = .078$. The mean SA for 1-computer teams was 15.54 (SD = 3.57), the mean SA for 2-computer teams was 20.18 (SD = 5.24), and the mean SA for 3-computer teams was 16.59 (SD = 2.02).

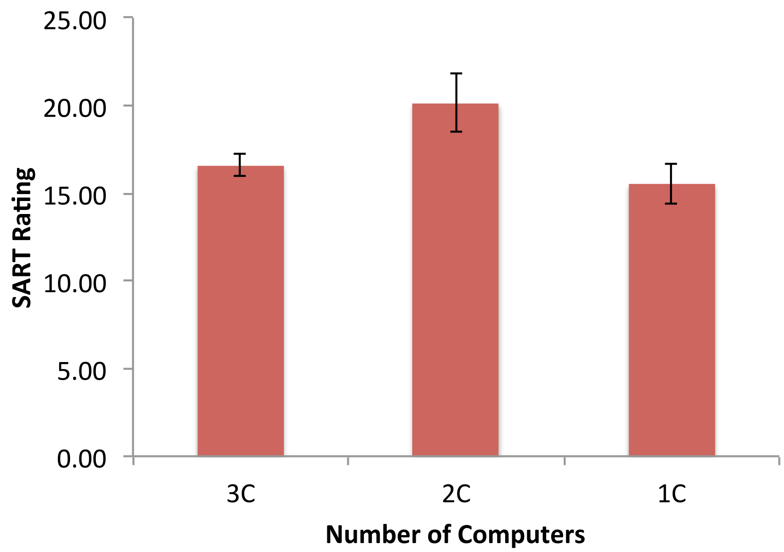


Figure 4.24 *SART ratings by resource allocation (mean values)*

As expected, the 1-computer teams reported having the least SA. The higher SA reported by the 2-computer teams than the 3-computer teams may indicate the benefit of having a clear facilitator role in the establishment of SA, which is particularly true for the 2-computer teams. As described earlier, the 1-computer and 3-computer conditions did not promote one clear facilitator.

The workload and SA results showed that these parameters were able to be influenced via resource allocation. However, ultimately there was not a strong indication of the significance of their impact on teamwork and performance, for this exercise. Workload (NASA-TLX) was not significantly correlated to Δ_{dec}, $r(34) = .153$, $p = .386$, nor Δ_{tot}, $r(34) =$

.175, p = .321. Also, SA (SART) was not significantly correlated to Δ_{dec}, $r(34)$ = .142, p = .422, nor Δ_{tot}, $r(34)$ = .181, p = .304.

4.3 Discussion

The purpose of Phase 2 was to determine the effect of resource allocation (number of computers) on teamwork and performance. As Phase 1 experiments showed the potential benefits of additional people and additional computers, the aim of Phase 2 was to further understand how this benefit could be accrued. Performance was not viewed as a simple overall number, but was viewed from the perspective of performance as an process of improving upon its taskwork, influenced by teamwork. As discussed in Chapters 3 and 4, the metrics Δ_{dec} and Δ_{tot} thus became the primary measures of interest.

Information as a "reduction of uncertainty" (Shannon & Weaver, 1949) was a theoretical premise for varying the numbers of computers. A computer, in this exercise, was the primary source for acquiring and generating information. Therefore, while keeping team size constant, it was expected that Δ_{dec} and Δ_{tot} would improve as a result of providing more sources of information (more computers). In addition, resource allocation in self-organizing teams was expected to affect team organization and communication interactions.

Overall, the results of Phase 2 indicated that the 3-computer teams demonstrated higher levels of Δ_{dec} and Δ_{tot} than 1-computer and 2-computer teams. And as expected, 3-computer teams demonstrated the most balanced communication interactions.

The 1-computer teams generally behaved in a way that was not expected, in that the 2 members without computers generally hovered over the one member with the computer, and attempted to ingest information in that manner. One assumption prior to Phase 2 was that the one computer condition may promote an organization having two facilitators and one analyst, in that the one analyst would be the one operating the computer and the other two members would interact with other interfaces available to them, such as the whiteboard or map. However, all ten teams adopted an organization in which they positioned themselves around the one computer for the majority of the time duration of the game.

Communication interactions varied, with 3-computer teams having the most balanced interactions, and the 1-computer teams having the least. The 1-computer teams also exhibited the most variance in their communication interactions, while the 3-computer teams had the least, showing that the 3-computer condition can be the most predictive structure for promoting balanced communications.

Again, teams were not given explicit instruction on how to behave, how to use various tools, and how to communicate. In the 1-computer condition, the large variance shows that people chose to do different kinds of things that may have resulted in the different communication interaction patterns. It is possible that if there had been more explicit instruction about roles or tool usage, the variance in 1-computer condition CIR could have been lower.

As Tollner's (2009) study showed, there can be a benefit of additional people in a task, be it that the people are able to contribute in a constructive way to the task. Phase 2 results showed that in the "Red/Blue" game, additional people are beneficial, but only when they are able to have access to their own source of information, manipulate and produce their own information, and monitor the team's performance by seeing how their actions affect the team. This result speaks strongly to the common interface that is provided through the computer, which offers an opportunity for every member something equally useful to do. This is made clearer when the Phase 2 results are viewed with respect to the Phase 1 individual participants. The normalized *d'* plot with the individual results are shown in Figure 4.25.

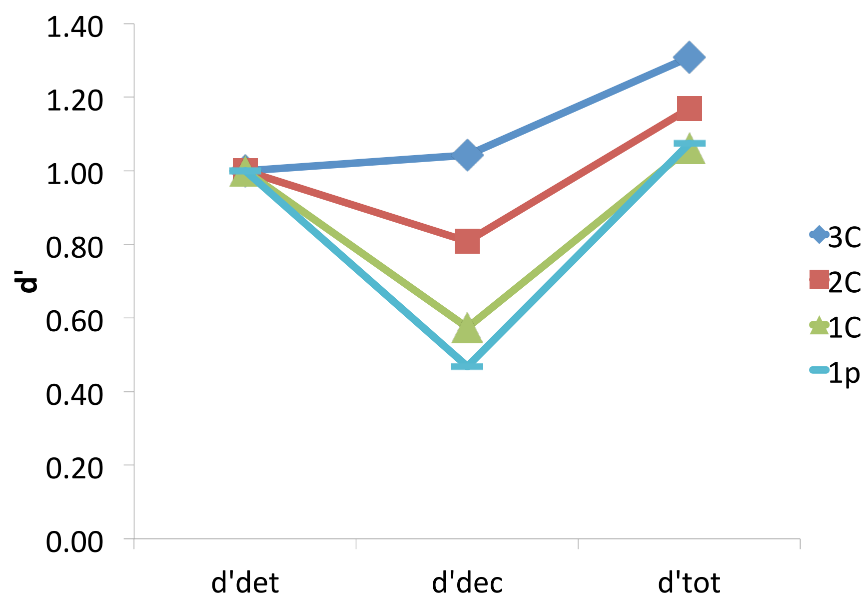


Figure 4.25 *Normalized Phase 2 results (d' values) by resource allocation with Phase 1 individual participant results (mean)*

Comparison of the 1-computer team with the individual participants shows the impact of two additional people without adding additional computers. There seems to be negligible improvement in d'_{dec}, which is not statistically significant, $t(18) = .488$, $p = .631$, and the two additional people provide essentially no gain in d'_{tot}.

Figure 4.25 illustrates the impact of adding people versus adding computers. Simply adding two people to the process had negligible effect.

However, when those people were added with sources of data (computers), in that they were given the ability to individually acquire, generate, and process their own information, then the decision process, teamwork, and performance were able to improve. The comparisons between the 3-computer teams and the individual participants clearly illustrate these insights.

5

Discussion and Conclusion

The series of experiments for this research made a unique contribution to the field in its study of the impact of resource allocation on teamwork and performance in self-organizing teams from a SDT framework.

The situation of the "Red/Blue" exercise, which was the basis for these experiments, is one in which teams begin with a high level of uncertainty. Then, using whatever resource or method is available to them, team members will seek to reduce their uncertainty, and the manner in which they do so reflects their effect on the team as either a drain or benefit. If we accept the definition of information as "reduction of uncertainty" (Shannon & Weaver, 1949) in this setting, people will act as active information seekers (Gibson, 1972). In relation to SDT, high uncertainty yields low awareness of the distinction between signal and noise, resulting in low sensitivity. With the addition of information, and the corresponding reduction in uncertainty, the disparity between the signal and noise curves will increase, thereby increasing sensitivity.

These experiments utilized self-organizing teams within a relatively short time duration. Members of such teams can come to be identified and committed to their teams (Chidambaram, Bostrom, & Wynne, 1990), and become a cohesive unit that can engage in collective action (Van De Van, Van De Vliert, & Oosterhof, 2003). As such, the intention of each team member is to contribute to the process by helping to reduce the team's uncertainty, but beneficial contribution will be promoted by the affordances of available resources, whether in the form of people or computers.

There were many insights provided by each phase of the experiments. Phase 1 showed that for this exercise, increasing the number of resources (people and computers) did not linearly improve performance.

In some cases, it degraded performance by adding extra noise to the process, which was reflected in the reduced teamwork that was seen in the largest groups. Phase 2 further explored the effect of increased computer resources while holding team size constant, and demonstrated the effect of providing an active information source (computer) to each member of the team. While this type of a team, in which every member has their own information source, did not promote a natural facilitator role, it promoted more balanced communication interactions, sense of shared information, which ultimately reflected in higher levels of teamwork. These results were compared to two other team concepts with one or two computers provided to the team. Teams with two computers promoted a natural facilitator, while teams with one computer resulted in somewhat unorganized, semi-chaotic organization patterns.

Performance was understood from the premise of information gain/loss and teamwork. From this understanding, the investigation was in evaluating the process from initial detection to final action phase, as an indication of how well the team had integrated their knowledge to ultimately choose which targets to take action against. This process was decomposed as a two-stage decision process: detection stage and decision stage. Teams showed different patterns of performance across the two stages. All teams whose sensitivity increased in the decision stage showed increased overall sensitivity. Some teams had some apparent loss of sensitivity at the decision stage from the detection stage, but were still able to have higher overall performance. And some teams lost sensitivity at the decision stage and resulted in lower overall sensitivity than they had at the detection stage.

In this sense, taskwork was able to be distinguished from teamwork, as d'_{det} emerged as a representation of a team's taskwork, while d'_{dec} emerged as a representation of a team's teamwork. The detection stage was understood to be the team building its knowledge base, by doing the common job of making detections, which was shown to be independent of factors such as team size. And it was further understood that it was the decision stage in which they had to perform as a team and teamwork properties emerged as being significant factors to performance in this stage. And while both stages, reflected by d'_{det} and d'_{dec}, were found to be significant contributors to overall performance, this research focused on the teamwork component of the process, acknowledging that each team could have had its own unique way of performing the taskwork component of the process.

5.1 Closed-loop system

This two-stage process can also be understood through basic control theory. A simple closed-loop control system, as shown in Figure 5.1, has the basic components of a reference signal, controller, plant, feedback, and output.

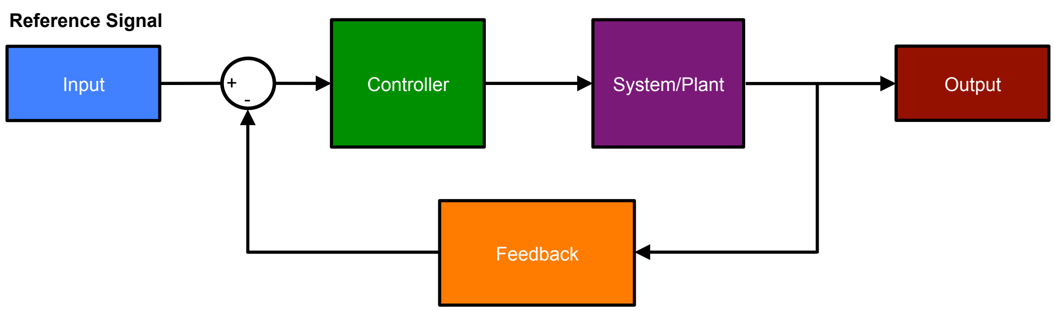


Figure 5.1 *Simple closed-loop control system*

Taking the above model as a framework to view the team process for the "Red/Blue" exercise, a similar construct can be established, shown in Figure 5.2.

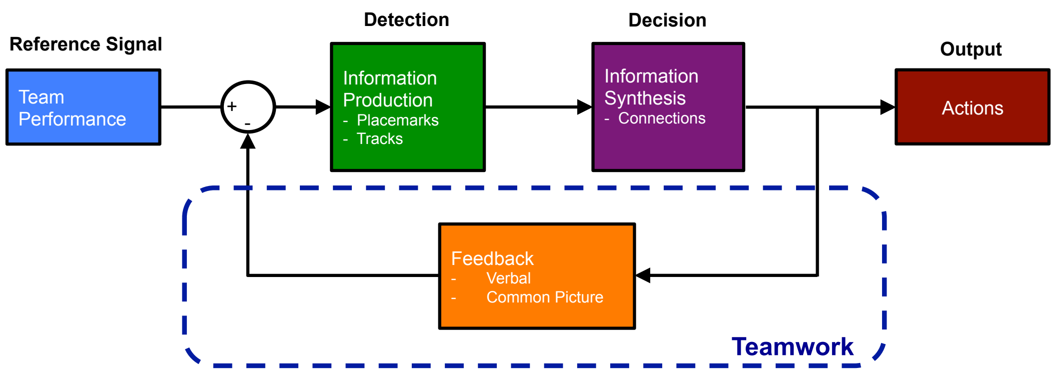


Figure 5.2 *Team performance and teamwork as a closed-loop control system*

In this model, the reference signal, the goal that teams are attempting to achieve, is team performance. From the input, the first step is the detection stage, when information production occurs. This represents the team's taskwork, and is detections in the form of creating placemarks and tracks. From these detections, the system or plant process occurs, as the decision stage. In this stage, information produced from the detection stage undergoes a synthesis process, where connections are made between placemarks, tracks, and confidence grows of the suspiciousness of certain sites.

In an open-loop concept of the system, the detections would feed into the decisions and result in actions without feedback. This was the case demonstrated with the individual participants. In the closed-loop concept shown above, there is a feedback loop, which is the teamwork component of the system. These thoughts are consistent with the taskwork and teamwork distinctions between the two stages, as the two-stage decision process was observed not as a linear progression between two stages, but a continuing decision development, as illustrated by the continuous loop in the system. The stages are distinct in the sense of detection and decision being two distinct states that teams undertake in through the course of a game. However, as the game progresses, the more team members are thinking about their decisions and developing their own picture of what may or may not be a suspicious site, they became better at distinguishing signal from noise and clearer on what their decisions are.

5.2 Delta

Phase 1 results, in light of the closed-loop system model, provide insights into a "delta" of performance (measured by d') as distinctions between a team's taskwork (d'_{det}) and its teamwork (d'_{dec}) began to emerge. It was found that the difference between sensitivity at decision and detection (Δ_{dec}) was significantly correlated with the difference of sensitivity between total and detection (Δ_{tot}). Patterns of performance emerged, as well. As Δ_{dec} increased (less negative) there tended to be an increase in Δ_{tot} (net increase in sensitivity). And when d'_{dec} exceeded d'_{det}, d'_{tot} always exceeded the sensitivity at detection (high Δ_{tot}). On the other hand, when Δ_{tot} was low, it suggested that there was some sort of interference.

Phase 2 extended these results by showing that the number of computers was positively correlated with both Δ_{dec} and Δ_{tot}. And since high Δ_{dec} was correlated with higher performance relative to the team's taskwork (Δ_{tot}), it stood to reason that allowing each member of the team to have their own computer to process their own information would result in consistently higher scores.

5.3 Effect of Resources

Team size literature supports the idea that, while other team members may contribute to team performance, they also have the potential to be detractors and suppress team performance (Amason & Sapienza, 1997; Shaw, 1981; Smith et al., 1994; Salas, et al., 1999; Guzzo, 1988; Guzzo & Shea, 1992; Hackman, 1990; Horwitz, 2005). Within the information theory and ecological psychology (Shannon & Weaver, 1949; Gibson, 1972) framework, there are many thoughts regarding how people can be detractors. People, as active information seekers, will work to resolve their uncertainty. If they do not have their own source of information, they will seek out information from others, and in so doing, could detract from those people's performance. If they are not facilitating a common picture amongst teammates, then their own seeking of information has no benefit to the team, and is in fact a drain. For example, in the 1-computer condition, this could be reflected in an imbalance of communication interactions, as those without computers communicate solely with the person with a computer, and not with each other.

A result regarding communication was that balanced communication interactions was a necessary but not sufficient condition for teamwork. As Phase 2 showed, every member of the team having his own computer did establish the conditions to facilitate balanced communication. The results do not conclusively indicate causality, in that teamwork is not caused by balanced interaction, nor is teamwork caused by everyone having their own computer. But the results do indicate that with those factors in place, the conditions promote an increased likelihood of teamwork.

Each person being able to be their own information producer because they have their own computer reduces the potential of interference, increases the number of targets that are considered, reduces groupthink that is not based on data, and drives a more data-driven approach than, perhaps, a hunch approach. And when everyone has his/her own computer, there was a flatter communication structure.

These conditions can be thought to facilitate the development of a shared mental map, where every team member was able to contribute to the solution, and see their contribution with respect to the team's overall solution. This can be contrasted with the condition where the interface was just one computer, and it was harder for the team to establish and maintain a shared mental map. The differentiations between the 1, 2, and 3-computer conditions can be recast in the sense that differences occur when people are given an opportunity to make their unique contribution.

The individual participants served as the baseline model for these comparisons. The gain, or lack of gain, in sensitivity as resources (people and computers) increased showed the effect of additional resources on teamwork and performance. The 1-computer team results showed that simply adding two people to solve the problem did not result in any significant difference in performance as compared to the 1-computer condition. To provide further comparisons, a look at the 1-person and 6-person results from Phase 1, with the Phase 2 results, is presented in Figure 5.3.

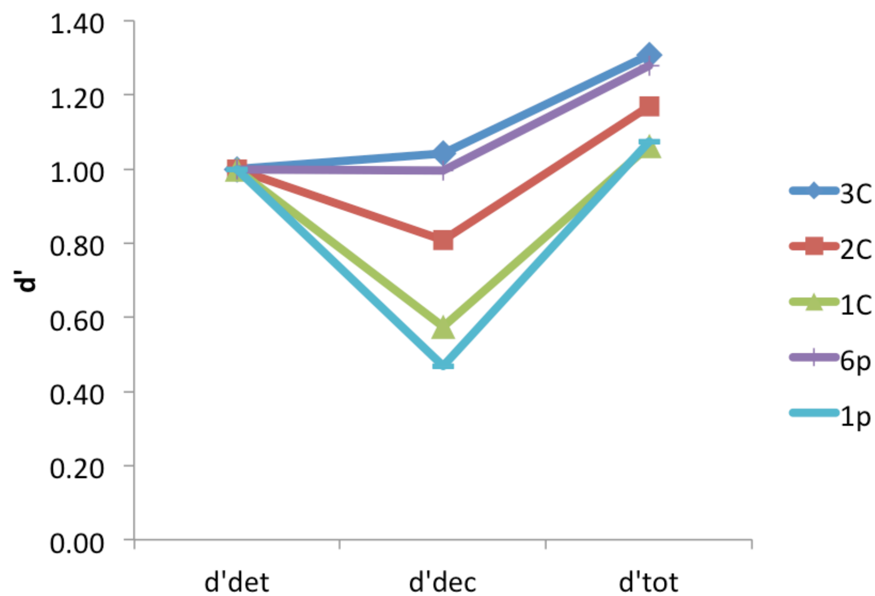


Figure 5.3 *Normalized Phase 2 results (d' values) by resource allocation with Phase 1 individual participants and 6-person team results (mean values)*

With the 6-person teams, a computer was added, and with the indications provided by Phase 2, it could be imagined that an additional computer would have provided a pure advantage. However, while a computer and a person to operate the computer were added, two additional people without computers were added as well, and they perhaps overwhelmed the advantage of the additional computer, as shown in Figure 5.3.

This is conjecture at this point, as it cannot be predicted what would happen if there were six people with six computers, or other combinations or people and computers, with this particular task, or a more complicated task. But these comparisons do give indications for future work.

5.4 Practical Implications

While the results of this research are based on a specific ISR simulation, and although the "Red/Blue" exercise is not a "real-life" scenario, it does emulate a command and control setting in which a team self-organizes and works together to solve a realistic problem. Thus, it gives rise to many practical implications regarding teamwork and potential topics for future work.

The issue of resource allocation in naturally forming teams and how to promote teamwork is an issue that exists across many fields, including classrooms and company organizations. The issue can be furthered explored from the perspective of how people interact with information, how certain interfaces promote different types of interactions, and how ultimately, these factors can be brought together to understand their decision making and performance.

The results provide implications regarding team training. Factors such as role training, personality, or trust building, while they cannot be discounted, perhaps become secondary factors to training from the perspective of information.

Other implications extend to the issue of leadership, and the type of structures that are imposed on organizations. To achieve successful teamwork, the results indicate the need to allow more transparency, allowing team members to see how what they do impacts the overall performance, and how they can adapt their own task performance to affect the overall team performance.

5.5 Future Work

It is hypothesized that two facilitating mechanisms of teamwork were (1) every team member having the ability to generate and process their own information, as well as (2) every team member having the ability to dynamically see their performance with respect to the overall performance of the team, thus allowing them to make adjustments to their own behavior to better assist the team.

These two conditions are made available by the particular tool used in this research. The first mechanism was demonstrated through the Phase 2 experiments. One potential area of future work will be to evaluate the second mechanism by shutting off the collaboration feature between computers for both the 3-computer condition and the 2-computer condition. This would allow the evaluation of the benefit of an interface that allows each person to see the placemarks/nominations/tracks that are being generated by other team members. With the 2-computer condition, the value of a facilitator could be enhanced. That person could function as the memory for the group, the one who generates the common map for the team. Without collaboration, it will be difficult for the people with computers to know what their performance on the total effect is. This will be especially difficult if collaboration is turned off and they are not allowed to communicate, which is another variable to consider for future work. Again, the value of the non-computer person is enhanced, as they would be the one to facilitate the communication, being the conduit to letting those with computers know their influence on team performance.

In this condition, for both the 2-computer and 3-computer conditions, it would be expected that there would be greater differentiation of workload, SA, and roles between the conditions. When everyone is not able to view all of the data, SA could be differentiated in a distinctive way between those with computers and those without. Workload may increase as well, as it would take more effort to establish and maintain the common picture. It could also be found that without the collaboration feature, the necessary and sufficient condition for teamwork is not the 3-computer condition. In the non-collaborative condition, establishing a shared mental model would require much effort. It would be predicted that, if there is going to be teamwork, someone would have to serve that function to generate the shared mental model. It would be less likely in the 3-computer condition than the 2-computer condition because a resource (a person) would need to be dedicated to that process.

Results from the exploration described above would give further implications to the research. It would suggest that in addition to providing each person their own information source, there is a need to allow some method for people to monitor their own performance with respect to the team's performance. If this cannot be established through the team member's computer interface, there would need to be other means to do so. This can take place through people, who are resources, or visual spatial representations, that are separate from the individual computer. These results would continue to be consistent with the premise of control theory and the role of feedback in the formation of teamwork. A means of feedback would be when individual contributors of a team are able to see what they are contributing and its influence on the team. In this sense, teamwork is transparency into knowing a person's influence and its effect on team performance, and in the opposite case, breakdown in teamwork can occur when any of the dyadic (or other) relationships within the team are not transparent and the person has no other way of compensating for the loss of information. Team size can be viewed from this perspective, as more members are added to a team beyond the optimal size for the task, the greater the chance for information loss.

Such results could provide further insights into team size. It may not be necessarily about the number of people, personalities within a team, or role identifications within the team, but it may be more about the common perception and awareness of team performance. The evaluation would be about who on the team knows what the true team performance is, and whether they can see how their own performance is influencing the team's performance. If this principle would hold true, team size or role identification may simply be artifacts of how the team was configured with respect to information.

Other areas for future work could involve the nature of the task itself. This research utilized an exercise in which one person with one computer could perform well. Future work could incorporate a more difficult exercise in which one person cannot do as well as teams. Within the "Red/Blue" exercise framework, the longer version of the game could be an option to increase difficulty.

Other future work regarding team organization could aim to increase the measurement of the dynamics within the team as team size is varied. More sensitive measurements for structure and communication could evaluate communication pattern structures. In measuring structure, the effect of role definitions could be measured. In a hierarchical structure, for example, the coordinator could be the common reference point for the team. By defining the role, a potential measurement would be to compare the *d'* of the coordinator to the *d'* of the team. With a true facilitator/coordinator, results could show that the *d'* of the coordinator is in fact the *d'* of the team.

Another topic of future work could be to better establish the contributions of additional people versus additional computers to the teamwork process. As results indicated from this research, for 3-person teams in the "Red/Blue" exercise considered here, additional computers had greater influence on teamwork and performance. 6-person results compared with the 3-computer results indicated that the additional people and computer didn't make much of a difference. However, this result alone was not conclusive in assessing the inner dynamic of additional people and computers, and further research would be required to accurately segregate those factors.

Lastly, more detailed modeling of the team SDT decision process could be a topic for future work. The SDT framework adopted for this research was a first-level, direct paradigm, which from the principle of Occam's razor, was appropriate to provide the necessary insights into the process. However, it is acknowledged that SDT was originally developed for individuals. Sorkin et al. (2001) have done some significant modeling of team performance in an SDT framework, but overall, this is still a very underdeveloped area. Future work could include methods to measure individual performance, and potentially evaluate how individual ROC's merge together in some sort of mathematical structure as a function of team size or resource allocation to represent the team ROC.

5.6 Conclusion

Results from Phase 1 showed that one person and one computer was a sufficient condition for good performance in this game. However, the results also showed that teams (addition of people) have the potential to provide greater consistency in performance. This consistency was observed as the result of teamwork, which resulted from effective acquiring, synthesizing, and sharing of information. Phase 1 showed that the "Red/Blue" exercise could be decomposed into a two-stage decision process, in which the team's taskwork (detection stage) and teamwork (decision stage) components could be distinguished. The correlations between Δ_{dec} and Δ_{tot}, and the correlations between Δ_{dec} and the communication interaction ratio, allowed an understanding of teamwork emerging as a function of team members' interactions with one another.

Phase 2 focused on the teamwork component of the process, in how teams work together to synthesize their information and make decisions. Performance, in this context, was viewed as d'_{tot} relative to d'_{det}, from the premise of information gain/loss and teamwork. Observations were made of the process from the detection stage to the decision stage, as an indication of how well the team had integrated their knowledge to ultimately decide which sites to take action against. This was reflected in Δ_{dec} and Δ_{tot}, and the investigation was regarding the factors that influenced how teams could improve upon their taskwork. To that extent, where appropriate, the results were normalized to d'_{det}. Again, this was not to diminish the effect of taskwork performance in its effect toward overall performance, but the normalization emphasized the internal dynamics of the teamwork component. By normalizing each team to its own d'_{det}, individual team differences in task performance were able to be subtracted out and allowed a focus on the decision stage performance in visualizing an isolated look at the effect of teamwork on performance, and the factors that most influence this teamwork. This was consistent with the approach of this research. Participants were randomly selected and recruited to participate in this game, as they are, not necessarily trained in ISR technologies or skills of network discovery. They were randomly assigned to teams and allowed to self-organize, which allowed the evaluation of their naturalistic behaviors: how they worked together, how they sustained or lost information, and how they made decisions as a team.

For the "Red/Blue" exercise, with 3-person teams, the results showed that everyone having their own computer is possibly the necessary and sufficient condition to produce teamwork. The results showed that additional people are beneficial, but only when they are able to have access to their own source of information, manipulate and produce their own information, and monitor the team's performance by seeing how their actions affect the team.

In a more general sense, in designing high performing teams, the goal then could be to facilitate each person's ability to generate and share their own information. This can be accomplished by giving each team member their own computer, or this can be accomplished through other means, such as having an effective facilitator. However, simply having a facilitator role is not the solution, because in some cases, the one without an information source can become a drain on the team. The unique characteristic of high performing teams where not everyone has their own computer is when the facilitator is not a drain and assists in developing the common picture.

The results of this research have the advantage of potentially recommending the least restrictive way of setting up teams. An example of this is a situation in which only one computer was available, and these results would suggest that only one person be put on the job, instead of having extra people to look over the shoulder of the person operating the computer. Compared to the individual participants, within 3-person teams, the value added of people and value added of resources were able to be seen.

Many of the topics that emerged through this research are relatively new areas of inquiry. The results revealed the impact of resources on teamwork and performance within self-organizing teams in a semi-realistic setting. The impact of people as resources and the impact of computers as resources were examined. In particular, the impact of the computer as an information source for each member of a team was seen as the primary driver for the emergence of teamwork.

The use of teams in complex decision making scenarios is an increasingly important element within many different types of organizations. As teamwork is recognized as and continues to be a significant factor in influencing overall performance, the resources and the allocation of those resources that promote effective information flow and teamwork should continue to be important factors to consider for team formation activities and future team research.

6

Appendices

Appendix A: Individual Participant (1p) Descriptives

Table A.1 *Individual Participant (1p) Descriptives*

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Total # PM	10	7.000	18.000	12.800	3.938	15.511
Hdet	10	2.000	8.000	4.600	2.319	5.378
FAdet	10	2.000	11.000	8.200	2.741	7.511
P(H)det	10	0.200	0.800	0.460	0.232	0.054
P(FA)det	10	0.011	0.058	0.043	0.014	0.000
d'det	10	0.778	2.568	1.630	0.649	0.422
Total # Actions	10	3.000	11.000	6.900	2.923	8.544
Hdec	10	1.000	7.000	3.500	2.014	4.056
FAdec	10	0.000	8.000	3.400	2.633	6.933
P(H)dec	10	0.500	0.938	0.741	0.152	0.023
P(FA)dec	10	0.045	0.889	0.421	0.255	0.065
d'dec	10	-0.546	2.758	0.957	1.024	1.048
d'dec (norm)	10	-0.659	1.315	0.483	0.526	0.277
Htot	10	1.000	7.000	3.500	2.014	4.056
FAtot	10	0.000	8.000	3.400	2.633	6.933
P(H)tot	10	0.100	0.700	0.350	0.201	0.041
P(FA)tot	10	0.003	0.042	0.019	0.013	0.000
d'tot	10	0.868	3.044	1.752	0.772	0.596
d'tot (norm)	10	0.867	1.451	1.074	0.154	0.024
Delta dec	10	-1.832	0.661	-0.673	0.683	0.466
Delta tot	10	-0.286	0.947	0.121	0.323	0.105
NASA-TLX	9	3.500	5.500	4.456	0.675	0.455
SART	10	10.000	21.000	14.700	3.743	14.011

Appendix B: 4-Person Team (4p) Descriptives

Table A.2 *4-Person Team Descriptives*

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Total # PM	6	14.000	27.000	21.830	4.535	20.567
Hdet	6	5.000	7.000	5.830	0.753	0.567
FAdet	6	9.000	22.000	16.000	4.336	18.800
P(H)det	6	0.500	0.700	0.583	0.075	0.006
P(FA)det	6	0.047	0.116	0.084	0.023	0.001
d'det	6	1.196	1.837	1.606	0.217	0.047
Total # Actions	6	6.000	11.000	9.000	2.000	4.000
Hdec	6	3.000	5.000	3.670	0.816	0.667
FAdec	6	3.000	7.000	5.330	1.633	2.667
P(H)dec	6	0.429	0.833	0.638	0.161	0.026
P(FA)dec	6	0.222	0.438	0.339	0.082	0.007
d'dec	6	0.157	1.446	0.810	0.489	0.239
d'dec (norm)	6	0.097	1.209	0.538	0.397	0.157
Htot	6	3.000	5.000	3.670	0.816	0.667
FAtot	6	3.000	7.000	5.330	1.633	2.667
P(H)tot	6	0.300	0.500	0.367	0.082	0.007
P(FA)tot	6	0.016	0.037	0.028	0.009	0.000
d'tot	6	1.264	1.938	1.579	0.218	0.048
d'tot (norm)	6	0.775	1.341	1.002	0.210	0.044
Delta dec	6	-1.473	0.250	-0.796	0.633	0.400
Delta tot	6	-0.367	0.408	-0.026	0.305	0.093
NASA-TLX	6	3.800	4.500	4.367	0.281	0.079
SART	6	16.300	24.000	18.883	2.834	8.034

Appendix C: 6-Person Team (6p) Descriptives

Table A.3 *6-Person Team Descriptives*

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Total # PM	12	12.000	37.000	20.500	6.626	43.909
Hdet	12	3.000	6.000	4.830	1.030	1.061
FAdet	12	7.000	31.000	15.670	6.272	39.333
P(H)det	12	0.300	0.600	0.483	0.103	0.011
P(FA)det	12	0.037	0.163	0.083	0.033	0.001
d'det	12	0.970	1.826	1.373	0.296	0.087
Total # Actions	12	3.000	15.000	7.330	4.075	16.606
Hdec	12	2.000	5.000	3.330	0.888	0.788
FAdec	12	0.000	11.000	4.000	3.931	15.455
P(H)dec	12	0.333	0.900	0.678	0.189	0.036
P(FA)dec	12	0.045	0.667	0.239	0.211	0.044
d'dec	12	0.122	2.365	1.375	0.686	0.471
d'dec (norm)	12	0.099	1.793	0.996	0.461	0.212
Htot	12	2.000	5.000	3.330	0.888	0.788
FAtot	12	0.000	11.000	4.000	3.931	15.455
P(H)tot	12	0.200	0.500	0.333	0.089	0.008
P(FA)tot	12	0.003	0.058	0.022	0.020	0.000
d'tot	12	0.829	2.537	1.754	0.508	0.258
d'tot (norm)	12	0.672	1.718	1.278	0.266	0.071
Delta dec	12	-1.113	1.046	0.002	0.609	0.371
Delta tot	12	-0.406	0.947	0.381	0.358	0.128
NASA-TLX	12	3.600	4.900	4.217	0.415	0.172
SART	12	12.200	22.000	17.817	3.061	9.371

Appendix D: 8-Person Team (8p) Descriptives

Table A.4 *8-Person Team Descriptives*

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Total # PM	2	14.000	42.000	28.000	19.799	392.000
Hdet	2	4.000	6.000	5.000	1.414	2.000
FAdet	2	10.000	36.000	23.000	18.385	338.000
P(H)det	2	0.400	0.600	0.500	0.141	0.020
P(FA)det	2	0.053	0.189	0.121	0.096	0.009
d'det	2	1.133	1.367	1.250	0.165	0.027
Total # Actions	2	6.000	18.000	12.000	8.485	72.000
Hdec	2	3.000	4.000	3.500	0.707	0.500
FAdec	2	3.000	14.000	8.500	7.778	60.500
P(H)dec	2	0.667	0.750	0.708	0.059	0.003
P(FA)dec	2	0.300	0.389	0.345	0.063	0.004
d'dec	2	0.713	1.199	0.956	0.344	0.118
d'dec (norm)	2	0.629	0.877	0.753	0.175	0.031
Htot	2	3.000	4.000	3.500	0.707	0.500
FAtot	2	3.000	14.000	8.500	7.778	60.500
P(H)tot	2	0.300	0.400	0.350	0.071	0.005
P(FA)tot	2	0.016	0.074	0.045	0.041	0.002
d'tot	2	1.196	1.625	1.410	0.304	0.092
d'tot (norm)	2	1.055	1.189	1.122	0.095	0.009
Delta dec	2	-0.420	-0.168	-0.294	0.179	0.032
Delta tot	2	0.426	0.483	0.455	0.040	0.002
NASA-TLX	2	3.600	4.300	3.950	0.495	0.245
SART	2	20.100	20.300	20.200	0.141	0.020

Appendix E: 1-Computer Team (1C) Descriptives

Table A.5 *1-Computer Team Descriptives*

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Total # PM	10	7.000	21.000	14.200	4.022	16.178
Hdet	10	3.000	7.000	5.400	1.265	1.600
FAdet	10	1.000	14.000	8.800	3.910	15.289
P(H)det	10	0.300	0.700	0.540	0.126	0.016
P(FA)det	10	0.005	0.074	0.046	0.021	0.000
d'det	10	1.202	2.811	1.850	0.504	0.254
Total # Actions	10	4.000	13.000	7.000	2.539	6.444
Hdec	10	2.000	6.000	3.900	1.287	1.656
FAdec	10	0.000	8.000	3.100	2.378	5.656
P(H)dec	10	0.571	0.917	0.709	0.121	0.015
P(FA)dec	10	0.062	0.571	0.369	0.189	0.036
d'dec	10	0.114	2.502	1.006	0.798	0.638
d'dec (norm)	10	0.068	1.263	0.573	0.442	0.195
Htot	10	2.000	6.000	3.900	1.287	1.656
FAtot	10	0.000	8.000	3.100	2.378	5.656
P(H)tot	10	0.200	0.600	0.390	0.129	0.017
P(FA)tot	10	0.003	0.042	0.017	0.012	0.000
d'tot	10	1.308	2.811	1.933	0.542	0.294
d'tot (norm)	10	0.803	1.409	1.061	0.213	0.045
Delta dec	10	-2.377	0.521	-0.844	0.936	0.877
Delta tot	10	-0.503	0.810	0.083	0.404	0.163
NASA-TLX	10	3.600	4.400	3.970	0.263	0.069
SART	10	10.300	20.700	15.540	3.569	12.736
CIR	10	1.073	4.432	2.079	1.322	1.747

Appendix F: 2-Computer Team (2C) Descriptives

Table A.6 *2-Computer Team Descriptives*

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Total # PM	15	9.000	21.000	14.533	4.015	16.124
Hdet	15	3.000	7.000	5.000	1.254	1.571
FAdet	15	4.000	16.000	9.533	4.224	17.838
P(H)det	15	0.300	0.700	0.500	0.125	0.016
P(FA)det	15	0.021	0.084	0.050	0.022	0.000
d'det	15	0.853	2.462	1.681	0.450	0.202
Total # Actions	15	3.000	9.000	5.467	1.767	3.124
Hdec	15	1.000	5.000	3.467	1.246	1.552
FAdec	15	0.000	5.000	2.000	1.813	3.286
P(H)dec	15	0.333	0.900	0.674	0.189	0.036
P(FA)dec	15	0.062	0.625	0.218	0.150	0.023
d'dec	15	0.244	2.563	1.373	0.661	0.437
d'dec (norm)	15	0.286	1.323	0.808	0.346	0.120
Htot	15	1.000	5.000	3.467	1.246	1.552
FAtot	15	0.000	5.000	2.000	1.813	3.286
P(H)tot	15	0.100	0.500	0.347	0.125	0.016
P(FA)tot	15	0.003	0.026	0.011	0.009	0.000
d'tot	15	0.751	2.791	1.967	0.586	0.343
d'tot (norm)	15	0.880	1.440	1.169	0.192	0.037
Delta dec	15	-1.123	0.625	-0.308	0.564	0.318
Delta tot	15	-0.200	0.853	0.286	0.318	0.101
NASA-TLX	15	3.900	4.900	4.327	0.258	0.066
SART	15	13.800	31.700	20.180	5.245	27.507
CIR	8	1.047	2.896	1.690	0.597	0.357

Appendix G: 3-Computer Team (3C) Descriptives

Table A.7 *3-Computer Team Descriptives*

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Total # PM	9	9.000	42.000	19.444	9.787	95.778
Hdet	9	2.000	8.000	5.444	1.810	3.278
FAdet	9	4.000	34.000	14.000	8.689	75.500
P(H)det	9	0.200	0.800	0.544	0.181	0.033
P(FA)det	9	0.021	0.179	0.074	0.046	0.002
d'det	9	0.731	2.286	1.631	0.452	0.204
Total # Actions	9	4.000	9.000	6.444	1.810	3.278
Hdec	9	2.000	5.000	4.111	1.167	1.361
FAdec	9	0.000	5.000	2.333	1.658	2.750
P(H)dec	9	0.571	0.833	0.753	0.097	0.009
P(FA)dec	9	0.059	0.455	0.207	0.150	0.022
d'dec	9	0.789	2.154	1.633	0.507	0.257
d'dec (norm)	9	0.551	1.790	1.042	0.356	0.127
Htot	9	2.000	5.000	4.111	1.167	1.361
FAtot	9	0.000	5.000	2.333	1.658	2.750
P(H)tot	9	0.200	0.500	0.411	0.117	0.014
P(FA)tot	9	0.003	0.026	0.013	0.008	0.000
d'tot	9	1.096	2.791	2.076	0.464	0.215
d'tot (norm)	9	1.058	1.701	1.308	0.205	0.042
Delta dec	9	-0.757	0.944	0.002	0.512	0.262
Delta tot	9	0.098	0.838	0.446	0.230	0.053
NASA-TLX	9	3.500	5.200	4.289	0.556	0.309
SART	9	13.000	19.300	16.589	2.022	4.089
CIR	9	1.004	1.405	1.107	0.118	0.014

Appendix H: Results by *A'*

One alternate method to calculate performance within the context of Signal Detection Theory is the metric *A'*, which is the area under a one-point ROC curve that corresponds to a particular *P(H)* and *P(FA)* (Macmillan & Creelman, 1991). When there is only one point in the ROC space, it is possible that the point lies on many different ROCs. A very conservative approach would be the area represented by *I*, in Figure A.1, which represents the minimum area. *S* represents the area that cannot possibly be included.

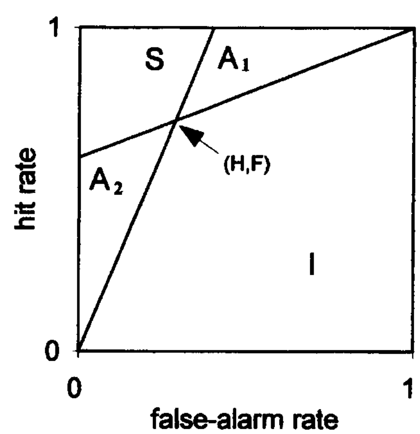


Figure A.1 *Calculation of area under a one-point ROC. A' is the minimum area plus one-half the sum of regions* A_1 *and* A_2 *(Macmillan & Creelman, 1996)*

Therefore, an estimate of the area was proposed by Pollack and Norman (1964) as a kind of average between the minimum and maximum performance.

$$A' = I + \frac{1}{2}(A_1 + A_2)$$

The normalized results are given below to demonstrate that similar conclusions can be made from *A'* results, as were made for *d'*.

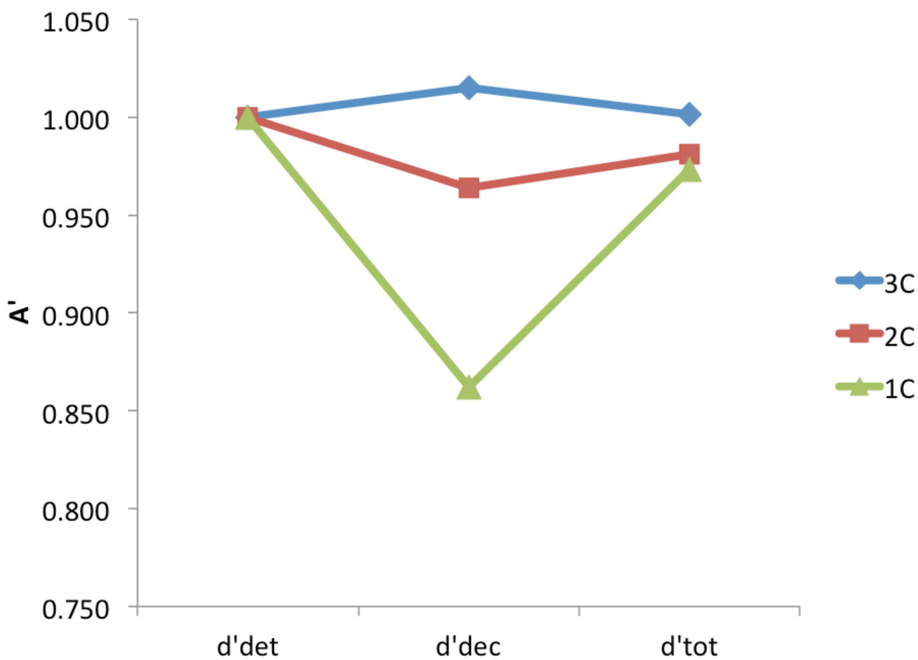


Figure A.2 *Normalized Phase 2 results (A' values) by resource allocation (mean)*

Appendix I: Command, Task Force Blue Latern

COMMAND, TASK FORCE BLUE LANTERN
FRAGO#: 206-09A

SITUATION.

1) UNION OF FORCES FOR DEMOCRACY AND DEVELOPMENT (UFDD) HAVE BEEN CONDUCTING ATTACKS ON UN PERSONNEL IVO ABECHE CHAD

2) JSOTF CERTAIN FURY WILL ARRIVE IN ABECHE TO PROVIDE SECURITY AND OPERATIONS SUPPORT

MISSION.

1) TRACK SUSPECT HOSTILES AND ID RED FACILITIES IVO ABECHE

CONCEPT OF THE OPERATION.

1) USE AIRBORNE ASSETS TO MONITOR ABECHE FOR OVERT HOSTILITIES AND DISCOVER RED NETWORK OF FACILITIES

PRIORITY INTELLIGENCE REQUIREMENTS (PIRS).

1) PROVIDE ESTIMATE OF RED FACILITIES TO JSOTF

2) IDENTIFY ADDITIONAL RED SITES FOR JSOTF ASSAULT PLANNING

3) IDENTIFY ADDITIONAL RED SITES FOR CONTINUED SURVEILLANCE

4) IDENTIFY MOST LIKELY RED HQ AS PART OF INTEL ESTIMATE

Figure A.3 *Command, Task Force Blue Lantern*

Appendix J: Scoring Matrix

	Red Facility	Red Activity	Gray Sites
Assault	+4	0	-2
Surveil	+2	+1	-1
No Action	-1	0	0

Figure A.4 *Scoring matrix*

Appendix K: NASA-TLX and SART Questionnaires

NASA Task Load Index (TLX)	Low 1	2	3	4	5	High 6	7
Mental Demand How mentally demanding was the task?							
Physical Demand How physically demanding was the task?							
Temporal Demand How hurried or rushed was the pace of the task?							
Performance How successful were you in accomplishing what you were asked to do?							
Effort How hard did you have to work to accomplish your level of performance?							
Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?							

	Situation Awareness Rating Technique (SART)	Low 1	2	3	4	5	High 6	7
	Instability of Situation How changeable is the situation? Is the situation highly unstable and likely to change suddenly (high), or is it very stable and straightforward (low)?							
	Variability of Situation How many variables are changing in the situation? Are there a large number of factors varying (high) or are there very few variables changing (low)?							
	Complexity of Situation How complicated is the situation? Is it complex with many interrelated components (high) or is it simple and straightforward (low)?							
	Arousal How aroused are you in the situation? Are you alert and ready for activity (high)? Or do you have a low degree of alertness (low)?							
	Spare Mental Capacity How much mental capacity do you have to spare in the situation? Do you have sufficient capacity to attend to many variables (high) or nothing to spare at all (low)?							
	Concentration How much are you concentrating on the situation? Are you bringing all your thoughts to bear (high) or is your attention elsewhere (low)?							
	Division of Attention How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (high) or focused on only one (low)?							
	Information Quantity How much information have you gained about the situation? Have you received and understood a great deal of knowledge (high) or very little (low)?							
	Information Quality How good is the information you have gained about the situation? Is the knowledge communicated very useful (high) or is it a new situation (low)?							
	Familiarity How familiar are you with the situation? Do you have a great deal of relevant experience (high) or is it a new situation?							

Appendix L: Team Workload, Teamwork, and Mutual Helpfulness Questionnaires

Please circle your assessment of each team member's role/roles and experienced workload (including your own)

Name	Role?				Experienced Workload
____________	Leader	Track Analyst	Graph Analyst	Other	1 2 3 4 5
____________	Leader	Track Analyst	Graph Analyst	Other	1 2 3 4 5
____________	Leader	Track Analyst	Graph Analyst	Other	1 2 3 4 5
____________	Leader	Track Analyst	Graph Analyst	Other	1 2 3 4 5

Teamwork Assessment

To what extent did team members provide relevant information to other team members, in a pro-active way, without that team member having to ask for it?

1 2 3 4 5

Never happened — Always happened

Were team members aware of one another's level of workload?

1 2 3 4 5

Totally unaware — Consistently aware

Did team members adjust individual task responsibilities to prevent overload?

1 2 3 4 5

No attempt made workload — Consistent efforts to redistribute when overload was detected

To what extent was the team's behavior coordinated?

1 2 3 4 5

Poor coordination — Excellent coordination

How congruent/similar were the leader's and the other team members' understanding of the mission?

1 2 3 4 5

Never in agreement — Completely in agreement

Mutual Helpfulness

In the questions below, rate how helpful each team member was to you by marking a position on the scale.

1. ____________

Hurt my job performance	Neither helped nor hurt my job performance	Helped my job performance

2. ____________

Hurt my job performance	Neither helped nor hurt my job performance	Helped my job performance

3. ____________

Hurt my job performance	Neither helped nor hurt my job performance	Helped my job performance

Team Workload Scale

Communication Demand

How critical was communication to team performance in the task?

1	2	3	4	5
Not critical		Somewhat critical		Critical

How hard did you have to work to request and/or transfer information?

1	2	3	4	5
No effort		Half of my effort		All of my effort

How much time was consumed in requesting information and/or responding to the requests?

1	2	3	4	5
None required		Half of the time		Constant

Monitoring Demand

How critical was monitoring team members to team performance in the task?

1	2	3	4	5
Not critical		Somewhat critical		Critical

How hard did you have to work to monitor others team members?

1	2	3	4	5
No effort		Half of my effort		All of my effort

How much time was consumed in monitoring other team members?

1	2	3	4	5
None required		Half of the time		Constant

Control Demand

How critical was correction of others to team performance in the task?

1	2	3	4	5
Not critical		Somewhat critical		Critical

How hard did you have to work to provide corrective feedback?

1	2	3	4	5
No effort		Half of my effort		All of my effort

How much time was consumed in correcting others?

1	2	3	4	5
None required		Half of the time		Constant

Coordination Demand

How critical was it to team performance to adjust your actions to coordinate with others?

1	2	3	4	5
Not critical		Somewhat critical		Critical

How hard did you have to work to adjust your actions to coordinate with others?

1	2	3	4	5
No effort		Half of my effort		All of my effort

How much time was consumed in adjusting your actions to improve team coordination?

1	2	3	4	5
None required		Half of the time		Constant

Leadership Demand

How critical was leadership to team performance in the task?

1	2	3	4	5
Not critical		Somewhat critical		Critical

How hard did you have to work to provide leadership?

1	2	3	4	5
No effort		Half of my effort		All of my effort

How much time was consumed in providing leadership?

1	2	3	4	5
None required		Half of the time		Constant

Appendix M: Training to Criteria

1) Go to a location in space (Manipulate map: zoom/pan)
Pan the map left click, hold, and move mouse
Zoom into the map, double left click to zoom into the map, or use the middle mouse scroll wheel to zoom in and out

2) List Layer Panel
- The left side panel contains a list of map layer items and messages
- Static Data (keep default settings – no need to modify)
- Sensor data layers
    - All off
    - EO tracks / dots
    - User Tracks / dots
- Messages and User Data
- EO dots and tracks are in light green
- User dots and tracks are in blue

3) Timeline
Move the timeline by left clicking and dragging the mouse left or right
- Red line: Current timeline time
- Data lines
    - Green – EO data
    - Blue – User data
    - Orange – Messages
- Data player controls (from left to right)
    - Snap to game start time
    - Step back
    - Play back
    - Stop
    - Play forward
    - Step forward
    - Snap to current game time

4) Tracks
Right click on an EO dot
Show track extent
- A track extent is comprised of a line that denotes the track history and squares denoting stops in the track history
- Right click on a track and select the option to show track extent
- To hide, right click on a track and select the option to hide track extent

Follow track
- Centers the spatial view on the current track point
- Right click on a track head and select the option to follow track
    - Note that the track history will be shown and highlighted in yellow

Unselect a track to follow

- Right click on the same track head (or track history) that you selected previously and select the option to stop following track
- Or use the stop following track button in the top left hand corner of Bluestreak

Highlight track

5) Messages

Double click "Messages"

- Left hand side panel
- Can fly to a message in space and in time

Right click on a message and select fly to option

Left click on a message to see details about a message

Hover over the message to see a summary about the message

6) How to create a site

Go to the track extent created above, pick a track end, and choose a building that's near it.

To create a placemark, right click and select option to "Make Placemark"

- Users can enter a name and a description for their placemark, along with a color (choose from 6 different colors – just note that two colors, pink and red are used for scoring)
- Can add notes
- Edit a created site

Hovering over a placemark will show and hide its label.

Left clicking on the placemark will display a pop up balloon with summary and detail information.

Double left clicking on a placemark zooms into the placemark

7) How to create a nomination

- The goal of a nomination is to nominate EO tracks that stop in a given spatial and time range, to find what tracks arrived or departed a particular area

Right click on the placemark created above

Make nomination

Select time range – note, making a large nomination time span will produce a lot of tracks

- The nomination dialog will display a timeline slider with two markers. The default position of the marker denotes your current timeline time. The markers denote the start time and end time of the nomination being created. By moving the left slider you change the start time of the nomination and by moving the right hand slider you change the end time of the nomination

Select spatial range – note, making a large spatial nomination will also produce a lot of tracks.

- The next slider is the radius slider. By default this is set to 50 meters.

- Note the results of the nomination
    - Track end points – will show up as an empty square
    - The tracks that were nominated are now user tracks that users can edit
    - Unconfirmed track stop to placemark connection
    - Show list panel organization
        - sites -> nominations -> track stops folder structure

Edit a nomination

- editing the nomination will re-run the nomination
- List layer panel update
- To edit a nomination region, right click on the nomination through the map or through the list layer panel, and then update the nomination parameters. This will re-run the nomination. Note that any track that was deleted will be re-nominated.

Delete a nomination

8) Working with Tracks

USER TRACKS – are copies of EO tracks that a user can edit. User tracks are represented by track positions and track stops

USER TRACK STOP TO PLACEMARK CONNECTION – track stop to placemark connection also have right click options that include "fly to" and deleted

USER TRACK MENU OPTION – Like EO tracks, by right clicking on a track position or track stop a user can select the "Follow track" option.

USER TRACK STOPS – user track stops are stops in a track's history and are represented by a blue square. Hovering over a track stop will display the track id and the time stamp or time range of the stop

TO VERIFY AND LINK TO PLACEMEMARK – *left click on the track stop and choose the option to link.* This produces what we call a rubber band line that can be used to connect a track stop to a placemark. Just move or stretch the rubber band line to the placemark that you wish to link the stop to. This will change the open track stop to a closed one and make the dashed line a solid one

WHERE DID THE BAD GUYS GO – when tracking back tracks that have left the region, go to their endpoint in time ">|" and then go out of track. To go out of track, to determine where the track goes, play forward, or click the + button to do a step forward (the step rate is denoted by the data player step rate slider, default step rate is 1 second)

WHERE DID THE BAD GUYS COME FROM

If track looks like it is useful, right click and select "Add to Graph"

9) Graph view

- The graph panel is the graphical representation of placemark, user tracks, user track stops, and user track connections
- All the right click options from the map view can be found here.
- Like map view, circles represent placemarks, and squares represent track stops. The track stop to placemark connections are represented by a line with arrows and

a time label, which represent the track's direction of travel and arrival/departure time in relation to the placemark. The track stop to track stop connection is a line with arrows that represent direction of travel between track stops.

Graph panel spatial view
Graph panel team view (nodes can be moved and reorganized to the team's liking. Left click and drag nodes around)

Collapse track stops – will collapse track stops into the placemark that they are connected to
Show track stop labels – shows/hides track stop labels
Show track edge labels – shows/hides the time labels on track stop to placemark connections

References

Adams, J. A. (2007). Unmanned vehicle situation awareness: A path forward. *Human Systems Integration Symposium*.

Adelman, L., Miller, A. L., & Henderson, D. (2003). Using Brunswikian theory and a longitudinal design to study how hierarchical teams adapt to increasing levels of time pressure. *Acta psychological, 112(2),* 181-206.

Alberts, D. S., & Hayes, R. E. (2003). *Power to the Edge: Command ... Control ... in the Information Age,* CCRP, Washington, DC.

Amason, A. C., & Sapienza, H. J. (1997). The effects of top management team size and interaction norms on cognitive and affective conflict. *Journal of Management*, *23*, 495-516.

Arney, M., Cohen, B., & Medairy, B. (2004). Impact of Advanced Collaborative Architectures on Intelligence Analysis, *2004 IEEE Aerospace Conference Proceedings*.

Aubé, C., Rousseau, V., & Tremblay, S. (2011). Team Size and Quality of Group Experience: The More the Merrier? *Group dynamics: Theory, Research and Practice.*

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Billman, L., & Steinberg, M. (2007). Human System Performance Metrics for Evaluation of Mixed-Initiative Heterogeneous Autonomous Systems. *NIST Performance Metrics for Intelligent Systems Workshop*.

Bolstad, C. A., & Endsley, M. R. (1999). Shared mental models and shared displays: an empirical evaluation of team performance. *Proceedings of the human factors and ergonomics society annual meeting, 4(3),* 213-217.

Bolstad, C. A., & Endsley, M. R. (2000). The effect of task load and shared displays on team situation awareness. Paper presented at the *14th*

Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society, Marietta, GA.

Bowers, C. A., Braun, C. C., & Morgan, B. B. (1997). Team workload: Its meaning and measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement* (pp. 85-108). Mahwah, NJ: Lawrence Erlbaum.

Bowers, C. A., Urban, J. M., & Morgan, B. B. (1992). *The Study of Crew Coordination and Performance in Hierarchical Team Decision Making*, Technical Report No. 92-01, University of Central Florida, Team Performance Laboratory, Orlando, FL.

Bradner, E., & Mark, G. (2002). Effects of team size on participation, awareness, and technology choice in geographically distributed teams. In *36th Hawaii International Conference on System Sciences* (pp. 150–160), IEEE Computer Society.

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty and sex on group performance and member reactions. *Journal of Personality and Social Psychology*, *36,* 1224-1440.

Bunderson, J. S., & Sutcliffe, K. M. (2002). Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects. *Academy of Management Journal*, *45*, 875–893.

Burke, C. S., Salas, E., Wilson-Donnelly, K., & Priest, H. (2004). How to turn a team of experts into an expert medical team: Guidance from the aviation and military communities. *Quality and Safety in Health Care, 13,* 96–104.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology, 91,* 1189–1207.

Burns, C. M. & Hajdukieqicz, J. R. (2004). *Ecological Interface Design* (pp. 4-5). Boca Raton, FL: CRC Press.

Campion, M. A., Papper, E. M., & Medsker, G. J. (1996). Relations between work team characteristics and effectiveness: A replication and extension. *Personnel Psychology, 49,* 429–452.

Cannon-Bowers, J. A., & Salas, E. (1990). *Cognitive psychology and team training: Shared mental models in complex systems.* Symposium address presented at the annual meeting of the Society for Industrial and Organizational Psychology, Miami, FL.

Cannon-Bowers, J. A., Salas, E., & Converse, E. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Current issues in individual and group decision making* (pp. 221-246). Hillsdale, NJ: Lawrence Erlbaum.

Chidambaram, L., Bostrom, R. P., & Wynne, B. E. (1990). A Longitudinal Study of the Impact of Group Decision Support Systems on Group Development, *Journal of Management Information Systems, 7(3),* 1990.

Clark, S. E., Hori, A., Putnam, A., & Martin, T. P. (2000). Group Collaboration in Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1578-1588.

Cooke, N. J., Gorman, J. C., Duran, J. L., & Taylor, A. R. (2007). Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology: Applied, 13,* 146–157.

Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 83-106). Washington, DC: American Psychological Association.

Cummings, M. L., & Nehme, C. E. (2010). Modeling the impact of workload in network centric supervisory control settings. *Neurocognitive and physiological factors during high-tempo operations*, 23-40.

DeSanctis, G., & Jackson, B. (1994). Coordination of Information

Technology Management: Team-Based Structures and Computer-Based Communication Systems, *Journal of Management Information Systems, 10(4)*, 85–110.

Drucker, P. F. (1988). The Coming of the New Organization, *Harvard Business Review, 66(5)*, 45–53.

Endsley, M. R. (1993). Situation awareness and workload: Flip sides of the same coin. In R. S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 906-911), Columbus, OH: Department of Aviation, The Ohio State University.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998), A comparative analysis of SAGAT and SART for evaluations of situation awareness. *42nd Annual Meeting of the Human Factors & Ergonomics Society,* Chicago, IL.

Ensley, M. D., Pearson, A., & Pearce, C. L. (2003). Top management team process, shared leadership, and new venture performance: A theoretical model and research agenda. *Human Research Management Review, 13,* 329-346.

Entin, E. E., & Entin, E. B. (2001). Measures for evaluation of team processes and performance in experiments and exercises. *Proceedings of the 2001 Command and Control Research and Technology Symposium.*

Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human Factors, 41(2),* 312-325.

Entin, E. E., Serfaty, D., & Deckert, J. C. (1994) *Team adaptation and coordination training (TR-648-1).* Burlington, MA: ALPHATECH, Inc.

Fern, E. F. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research, 19,* 1-13.

Fincannon, T. D., Evans, A. W., Philips, E., Jentsch, F., & Keebler, J. (2009). The Influence of Team Size and Communication Modality on

Team Effectiveness with Unmanned Systems, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53*, 419-423.

Fiore, S. M., & Salas, E. (2004). Why we need team cognition. In: S. Fiore & E. Salas (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 235-48). Washington, DC: American Psychological Association.

Fiore, S. M., & Schooler, J. W. (2004). Process mapping and shared cognition: teamwork and the development of shared problem models. In: E. Salas & S. M. Fiore (Eds.), *Team cognition: understanding the factors that drive process and performance* (pp. 133-152). Washington, DC: American Psychological Association.

Forssell, D. G. (2008). Once around the loop: An interpretation of basic PCT. In *Perceptual Control Theory: science & applications - a book of readings.* Hayward, CA: Living Control Systems Publishing.

Galbraith, J. (1973). *Designing Complex Organizations*. Reading, MA: Addison-Wesley.

Gibson, J. J., & Gibson, E. (1972). The Senses as Information-Seeking Systems, *Times Literary Supplement*, 711.

Gibson, J. J. (1979), *The Ecological Approach to Visual Perception*, Boston, MA: Houghton-Mifflin.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Guastello, S. J. (2010). Nonlinear Dynamics of Team Performance and Adaptability in Emergency Response. *Human Factors: The Journal of the Human Factors and Ergonomics Society 2010, 52,* 162.

Guzzo, R. A. (1988). Financial incentives and their varying effects on productivity. In P. Whitney & R. B. Ochsman (Eds.), *Psychology and productivity* (pp. 81–92). New York, NY: Plenum Press.

Guzzo, R. A., & Shea, G. P. (1992). Group performance and intergroup relations in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 269-313). Palo Alto, CA: Consulting Psychologists Press.

Hackman, J. R. (Ed.). (1990). *Groups that work (and those that don' t).* San Francisco: Jossey-Bass.

Hancock, P. A., & Ganey, N. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. *Journal of Human Performance in Extreme Environments, 7(1),* 5–14.

Hare, A. P. (1981). Group size. *American Behavioral Scientist, 24*, 695-708.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 50(9)*, 904-908

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North-Holland, 139-183.

Hebb, D. O. (1955). Drives and the C.N.S. (Conceptual Nervous System). *Psychological Review, 62,* 243-254.

Heeger, D.J. (1997). *Signal detection theory*, http://www.white.stanford.edu/~heeger/psych202/readings/sdt handout.pdf

Hendy, K. C. (1995). Situation awareness and workload: Birds of a feather? *Situation Awareness: Limitations and Enhancement in the Aviation Environment*. Neuilly sur Seinne, FR: AGARD 21-1 – 22-7.

Henry, R. A. (1995). Improving group judgment accuracy: information sharing and determining the best member. *Organization of Behavioral Human Decision Processes, 62(2)*, 190-97.

Highsmith, J. (2004). *Agile Project Management: Creating Innovative Products*. Addison-Wesley, USA.

Hildebrand, G. A., Pharmer, J. A., & Weaver, M. D. (2003). Performance and usability evaluation of an automated task management concept prototype design. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting,* 1913-1917.

Hinsz, V. B. (1990). Cognitive and consensus processes in group recognition memory performance. *Journal of Personality and Social Psychology, 59 (4),* 705-718.

Hoda R., Noble J., & Marshall, S. (2010). Organizing self-organizing teams. *ICSE2010, ACM*, South Africa, 285–294.

Hollenbeck, J. R., DeRue D. S., & Guzzo, R. (2004). Bridging the gap between I/O research and HR practice: Improving team composition, team training, and team task design, *Human Resource Management 43(4),* 353-366.

Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., Colquitt, J. A., & Hedlund, J. (1998). Extending the multilevel theory of team decision making: Effects of feedback and experience in hierarchical teams. *Academy of Management Journal, 41(3).* 269-283.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise, *Journal of Applied Psychology 80 (2)*, 292-316.

Horwitz, S. K. (2005). The compositional impact of team diversity on performance: Theoretical considerations. *Human Resource Development Review, 4*, 219-245.

Humphrey, S. E., Hollenbeck, J. R., Meyer, C. J., & Ilgen, D.R. (2002). Hierarchical team decision making, *Research in Personnel and Human* Resources Management, (21), 175-213.

Ilgen, D. R., Major, D., Hollenbeck, J. R., & Sego, D. (1995). Raising an individual decision making model to the team level: A new research model and paradigm. In R. Guzzo & E. Salas (Eds.) *Team Effectiveness and Decision Making in Organizations* (pp. 113-148), Jossey-Bass, San Francisco.

Jehn, K. A., & Shah, P. P. (1997). Interpersonal relationships and task performance: An examination of mediating processes in friendship and acquaintance groups. *Journal of Personality and Social Psychology, 72,* 775–790.

Johnson, S. D., Suriya, C., Yoon, S. W., Berrett, J. V., & Fleur, J. L. (2002). Team development and group processes of virtual learning teams, *Computers & Education, 39(4),* 379-393.

Kang, M. (2010). A proposed improvement to the multilevel theory for hierarchical decision-making teams. *Journal of Management & Organization, 16(1)*, 151-167.

Kim, P. K. S., & Martin, J. S. (2009). *Team Jesus.* Maitland, FL: Xulon Press.

Klein, K. J., Ziegert, J. C., Knight, A. P., & Xiao, Y. (2006). Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative Science Quarterly, 51*, 590-621.

Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (pp. 333–375). London, England: Wiley.

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006) Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Social Psychology, 90,* 644 –651.

LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., Colquitt, J. A., & Ellis, A. (2002). Gender composition, situational strength, and team decision-making accuracy: A criterion decomposition approach. *Organizational Behavior and Human Decision Processes 88(1),* 445-475.

Levchuk, G. M., Yu, F. Levchuk, Y., & Pattipati, K. R. (2004). Networks of Decision-Making and Communicating Agents: A New Methodology for Design and Evaluation of Organizational Strategies and Heterarchical Structures. *Proceedings of the 2004 International Command and Control Research and Technology Symposium*, San Diego, CA.

Lewelling, T. A., & Nissen, M. E. (2007). Hypothesis Testing of Edge Organizations: Laboratory Experimentation using the ELICIT Multiplayer Intelligence Game. *Proceedings of the 12th International Command and Control Research and Technology Symposium*, Newport, RI.

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology, 88*, 587–604.

Lipnack, J., & Stamps, J. (2000). *Virtual Teams: People Working Across Boundaries with Technology (2nd ed.),* Wiley, New York.

McCue, C. (2005). Data Mining and Predictive Analytics: Battlespace Awareness for the War on Terrorism, *Defense Intelligence Journal 13 (1&2),* 47-63.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Lessons from complex environments. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9-45), San Francisco: Jossey-Bass.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and bias. *Psychonomic Bulletin & Review, 3,* 164-170.

Majalian, K. A., Kleinman, D. L., & Serfaty, D. (1992). The effect of team size on team coordination. *IEEE International Conference on Systems,* *Man and Cybernetics*, 1, 880-886.

Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team interaction training for team adaptation to novel environments. *Journal of Applied Psychology, 85,* 971–986.

Martin, D., & McEver, J. (2008). Metrics, Analysis and Methods for the Exploitation of ELICIT Experimental Data, *Proceedings of the 13th International Command and Control Research and Technology Symposium*, Seattle, WA.

Martins, L. L., Gilson, L. L., & Maynard, M. T. (2004). Virtual Teams? What do we know and where do we go from here? *Journal of Management, 30,* 805–835.

Massey, A. P. (1991). *Group Problem Definition: The Impact of Network Structuring Techniques on Group Effectiveness and Interaction*, Unpublished doctoral dissertation, Rensselaer Polytechnic Institute, Troy NY.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85,* 273-283.

Mesmer-Magnus, J., & DeChurch, L. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94,* 535-546.

Moreland, R. (1999). Transactive memory: Learning who knows what in work groups and organizations. In L. Thompson, J. Levine, & D. Messick (Eds.), *Shared cognition on organizations* (pp. 3-31). Hillsdale, NJ: Lawrence Erlbaum.

O'Neil, H. F., Baker, E. L., & Kazlauskas, E. J. (1992). Assessment of team performance. In: R. W. Sweezy & E. Salas (Eds.) *Teams: Their training and performance*, (pp. 153-175), Ablex Publishing, Norwood NJ.

Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In: G. Klein, J. Orasanu, R. Calderwood, C. E. Zsambok (Eds.) *Decision making in action: models and methods* (pp. 327-45). Norwood, NJ: Ablex.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30(3),* 286-297.

Pearsall, M. J., & Ellis, A. P. J. (2010). Building the infrastructure: The effects of role identification behaviors on team cognition development and performance. *Journal of Applied Psychology, 95(1),* 192-200.

Peterson R. S., Owen, P. D., Tetlock, P. E., Fan, E. T., & Martorana, P. (1998). Group dynamics in top management teams: groupthink, vigilance, and alternative models of organizational failure and success. *Organizational Behavior and Human Decision Processes, 73,* 272–305.

Pett, M., Lackey, N. R., & Sullivan, J. J. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. London: Sage.

Phillips, J. M. (1999). Antecedents of leader utilization of staff input in decision-making teams, *Organizational Behavior and Human Decision Processes 77(3),* 215-242.

Phillips J. M. (2002). Antecedents and consequences of procedural justice perceptions in hierarchical decision-making teams. *Small Group Research, 33,* 32–6.

Powers, W. T. (1973). Feedback: beyond behaviorism, *Science, 179 (4071),* 351-356.

Priest, H. A., Stagl, K. C., Klein, C., & Salas, E. (2006). Virtual Teams: Creating Context for Distributed Teamwork. In C. Bowers, E. Salas, & F. Jentsch (Eds.), *Creating High-Tech Teams: Practical Guidance on Work Performance and Technology* (pp. 185–212), American Psychological

Association, Washington, DC.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology, 53 (1),* 61-86.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On Teams, Teamwork, and Team Performance: Discoveries and Developments Human Factors: *The Journal of the Human Factors and Ergonomics Society, 50,* 540.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance.* Norwood, NJ: Ablex.

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

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine, 78(5 Suppl),* B77-B85.

Salas, E., Rozell, D., Driskell, J.D., & Mullen, B. (1999). The effect of team building on performance: An integration. *Small Group Research, 30,* 309–329.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a “big five” in teamwork? *Small Group Research, 36,* 555-599.

Salas, E., Stagl, K. C., Burke, C. S., & Goodwin, G. F. (2007). Fostering team effectiveness in organizations: toward an integrative theoretical framework of team performance. In J. W. Shuart, W. Spaulding, & J. Poland (Eds.), *Modeling complex systems: motivation, cognition and social processes.* Nebraska Symposium on Motivation. Lincoln, NE: University of Nebraska Press.

Scerri, P., Owens, S., Sycara, L., & Lewis, M. (2010). User evaluation of a GUI for controlling an autonomous persistent surveillance team. *Proceedings of SPIE,* 7694.

Schraagen, J. M., Veld M. H., & De Koning, L. (2010). Information Sharing During Crisis Management in Hierarchical vs. Network Teams. *Journal of Contingencies and Crisis Management, 18(2),* 117-127.

Selcon, S. J., Taylor, R. M., & Koritsas, E. (1991). Workload or situational awareness? TLX vs. SART for aerospace systems design evaluation, in *Proceeedings of the Human Factors Society's 35th Annual Meeting*, Santa Monica, CA, USA: Human Factors Society, 62-66.

Serfaty, D., Entin, E. E., & Johnston, J. (1998). Team Adaptation and Coordination Training. In J. A. Cannon-Bowers & E. Salas (Eds.) *Decision Making Under Stress: Implications for Training and Simulation.* Washington D.C.: APA Press.

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

Shaw, D. M. (1960). Size of share in task and motivation in work groups. *Sociometry, 1960(23),* 203-208.

Shaw, M. E. (1981). *Group dynamics: The psychology of small group behavior (3rd ed.).* New York: McGraw-Hill.

Sheridan, T. B. (1979). Mental workload in decision and control. *Decision and Control including the Symposium on Adaptive Processes, 18th IEEE Conference, 18,* 977-982.

Smith. K. G., Smith, K. A., Olian, J. D., Sims, H. P., O'Bannon, D. P. & Scully, J. A. (1994). Top management team demography and process: The role of social integration and communication, *Administrative Science Quarterly, 3Y, 4*, 12-438.

Solansky, S. T. (2008). Leadership Style and Team Processes in Self-Managed Teams. *Journal of Leadership & Organizational Studies, 14 (4),* 332-341.

Somech, A. (2006). The effects of leadership style and team process on performance and innovation in functionally heterogeneous teams. *Journal of Management, 32,* 132-15.

Sorkin, R. D. (1999). Spreadsheet signal detection. *Behavior Research Methods, Instruments, and Computers, 31,* 46–54.

Sorkin, R. D., Hays, C. J., & West, R. (2001). Signal-detection analysis of group decision making. *Psychological Review, 108(1),* 183.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48,* 1467–1478.

Stasser, G. & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology, 53,* 81–93.

Taggar, S., & Brown, T. C. (2001). Problem-solving team behaviors: Development and validation of BOS and a hierarchical factor structure. *Small Group Research, 32,* 698- 726.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401-409.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design, in *Situational Awareness in Aerospace Operations (AGARD-CP-478),* Neuilly Sur Seine, France: NATO-AGARD, 3/1-3/17.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory and Psychology, 4(4),* 525-547.

Tollner, A. M. (2009). *Individual and team susceptibility to change blindness.* Doctoral dissertation, University of Cincinnati.

Urban, J. M., Weaver, J. L., Bowers, C. A., & Rhodenizer, L. (1996). Effects of Workload and Structure on Team Processes and Performance: Implications for Complex Team Decision Making, *Human Factors, 38,* 300–310.

Van De Ven, G.S., Van De Vliert, E., & Oosterhof, A. (2003). Informational Dissimilarity and Organizational Citizenship Behavior: The Role of Intrateam Interdependence and Team Identification. *Academy of Management Journal, 46(6),* 715-727.

Verma, S., Kozon, T., Cheng, V., & Ballinger, D. (2007). Changes in roles/responsibilities of Air Traffic Control under precision taxiing. *IEEE/AIAA 26th Digital Avionics Systems Conference.*

Von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology, 5,* 520-526.

Walker, G. H., Stanton, N. A., Salmon, P., Jenkins, D., Monnan, S., & Handy, S. (2008). An Empirical Evaluation of Network Centric Team Organisation Versus Hierarchical Team Organisation, *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, 458–462.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). New York/ Berlin: Springer-Verlag.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance (3rd ed.).* Columbus, OH: Charles E. Merrill.

Wickens, T. D. (2002). *Elementary signal detection theory.* Oxford: Oxford University Press.

Won, J. C., Condon, G. R., Landon, B. R., Wang, A. R., & Hannon, D. J. (2011). Assessing team workload and situational awareness in an Intelligence, Surveillance, and Reconnaissance (ISR) simulation exercise. *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference,* 163-167.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330,* 686–688.

Xu, K., & Zhao, L. (2011). Individual Swift Trust and Cooperation in Emergency Rescue Team Members. *ICSSC*, 2011.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology, 18 (5),* 459-482.

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *Leadership Quarterly, 12,* 451-483.