Chapter 2 – RELATIONSHIP BETWEEN CULTURE, INFORMATION SHARING, SITUATION AWARENESS, AND TEAM PERFORMANCE

2.1 BACKGROUND

A five-nation, distributed team laboratory experiment was conducted in Bulgaria, The Netherlands, Norway, Sweden, and The United States and was referenced at the "Leader and Team Adaptability in Multinational Coalitions (LTAMC)" experiment. The goal was to investigate whether team composition (single or multi-nation) has an impact on information sharing, situation awareness and team performance. Specifically, a single experimental design was executed to build a baseline of responses from nationally homogenous teams from the participating countries. Then, that design was executed with nationally heterogeneous teams comprised of participants from the different participating countries. Team composition on culturally heterogeneous teams was randomized using a Latin Square design developed by Capt (BUL-N) Dr. Yantsislav Yanakiev. The experiment objective was to answer the following questions:

- What is the relationship among measures of team process and outcome in a distributed task environment where the members are not face-to-face and represent different nationalities?

- In a distributed task, are nationally homogeneous teams more effective at information sharing, developing situation awareness, and ultimately achieving better team performance outcomes than nationally heterogeneous teams?

- Can measures of cognitively-based cultural biases distinguish national teams from multi-national teams?

Team performance is, in part, the result of team members effectively performing their interdependent tasks. Information sharing and situation awareness (SA) are widely recognized to be important factors in achieving effective team performance outcomes. Effective information sharing has been linked to situation awareness which is a critical cognitive skill required for maintaining awareness, knowledge, and understanding of events in our immediate and future environment [7]. Teams that share task-related information are more likely to have better SA, make better decisions and have better overall performance outcomes than teams that do not exchange information ([8], [9], [10]). Likewise, we would expect effective information sharing and situation awareness to be especially important factors in a distributed team performance environment, but little research exists on this subject.

Fiore and colleagues [11] have proposed cognitive explanations for factors that would impact the effectiveness of distributed team members. When team members are co-located, they have the advantage of clear perceptual clues to help them discern meaning and intention from fellow team members' paralinguistic cues during face-to-face meetings [11], [12]. When team members become distributed, the rich visual, auditory, and social array of cues is lost and they must rely on less reliable technical forms of communication to discern meaning and intention [11]. Moreover, when team members are physically close to each other, the frequency of communication is greater than for team members who are more physically distant [13]. Consequently, the ability for efficient information sharing and the development of good situation awareness is compromised. As a result, efficient and effective team decision-making and team performance outcomes are likely to be compromised. Nevertheless, we expect to see some positive relationships among the dependent variables in this experiment. Furthermore, however, we expect that nationally homogeneous military teams will perform better than multi-national teams in a distributed environment because they share similar training experiences and expectations.

2.1.1 National Culture

Sutton et al. propose that culture can be construed as a cognitive framework within which personal experiences are interpreted, values are formed, and behaviour results, to include team behaviours [14], [15], [16]. The expectation is that cultural norms may determine what information can be shared up and down the chain of command or among peers, or how much information is needed to make a decision, among other things. Sutton and Pierce describe the construct of "cultural competence" as having a clear understanding of the dominant values and culturally based orientations (e.g., relationship orientation, time orientation) of other team members, and recognizing that the thoughts and behaviours of others are influenced by such cultural norms. Consequently, diversity in cultural competence, combined with the lack of natural environmental and social cues, may explain why the incidence of conflict within the team increases as teams become more distributed and multinational, (for review see [17]). Together, the increasing multinational and distributed nature of teams can significantly impair the ability of team members to perform well because they are less able to effectively engage in the communications necessary to enhance information sharing and situation awareness. Therefore, in this experiment we explored the degree of similarity in responses to the GlobeSmart® Commander cultural dimensions due to team composition (single nation vs. multi-nation).

2.1.2 Situation Awareness

Effective decision-making is a vital component to the success of any organization [7]. However, within the human factors community, there is a general belief that effective decision-making is facilitated by good situation awareness (SA) [9]. Although the definition of SA varies in broad terms, broadly, SA is a cognitive construct that refers to our awareness, knowledge, and understanding of events in our immediate and future environment [7]. Since World War I, researchers and practitioners have come to view SA as critical for accurate decision-making and performance in a variety of work domains such as air traffic control [18], nuclear power plant management [18], aviation [21], medicine [21], and driving [22]. While most of this research has focused primarily on understanding and developing systems to support individual operator SA, in reality, teams of operators perform work; teams of operators run nuclear power plants; crews of pilots fly planes; and groups of medical personnel treat patients [26]. Indeed, SA has been identified as an important component of the success of team efforts [8], [9], [10], and researchers are now focussing more of their attention on the study of shared or team SA.

An interesting trend occurring with teams (both civilian and military) and organizations in general is that they are becoming more culturally diverse and geographically distributed. To date, there is a paucity of research designed to examine the relationship between culture and situation awareness. However, there have been numerous studies developed to examine the relationship between culture and decision-making. For example, researchers have demonstrated cultural differences in planning and decision-making effort [27], judgment and decision-making (for review see [28]), confidence in decision-making, decision-making styles, risk-taking [23]. [24], [25], [26], [27], [28], [29], [30]. Interestingly, much of this research has focused primarily on describing national differences in various facets of decision-making with much less research aimed at examining the underlying mechanisms for these differences [28], [30]. Because SA is considered a vital component of decision-making, and little is known about the underlying mechanisms that explain cultural differences in decision-making, it seems possible that the observed cultural differences in decision-making might, at least in part, be due to cultural differences at the level of SA. Thus, a fruitful avenue for SA research specifically, and for decision-making in general, would be to examine the relationship between culture and SA.

As teams become more distributed [11], the ability to develop good SA becomes more difficult than for teams that are co-located. When teams are co-located, team members have the advantage of clear perceptual clues to help them discern meaning and intention from fellow team members' paralinguistic cues during face-to-face meetings [11],[12]. When teams become distributed, the rich visual, auditory, and social array of cues normally present in co-located teams is lost and team members must rely on less reliable technical forms of communication to discern meaning and intention [11],[13]. Moreover, when team members are physically close to each other, the frequency of communication is greater than for team members who are more physically distant [13]. In addition, as teams become more geographically distributed and multinational, the incidence of conflict within the team increases. Together, the increasing multinational and distributed nature of teams, as well as other characteristics (for further discussion see [13]), tax the ability of team members to engage in effective communication. Consequently, the ability for efficient information sharing and the development of good SA is compromised. As a result, efficient and effective team decision-making is likely to be compromised.

Both cultural diversity and distributed teams pose important challenges to the development of good team SA. This is a critically important issue to militaries around the world as these militaries increasingly perform within culturally diverse teams and become more geographically distributed. The challenge for researchers is to examine the extent to which these factors in particular and related factors (e.g., cognitive biases, trust), individually or in combination with each other, impact team decision-making. Armed with this knowledge, researchers can then assist in the development of new concept of operations and training programs to help alleviate the potential negative impacts that cultural diversity and distributed teamwork have on the development of good team SA and ultimately efficient team decision-making.

2.1.3 Team Performance

The understanding of a team performance perspective for developing effective organizations has been recognized in the research literature especially during the last two decades due to a number of reasons [31]. However, even if an interest for team performance exists in the research community and quite a large quantity of articles focus on team performance the understanding of team processes and especially the relation between team processes and outcome still leaves practitioners with a number of unanswered questions.

Team performance is a complex domain to study. Not only are there often many participants involved (compared to studies of individuals) but it can also be difficult to identify and measure relevant team performance processes or outcome. There is a need to integrate results and knowledge provided by different research domains within the team performance field. Models and important dimensions for team work need to be integrated with research about the mechanisms that can explain how the interaction between team members affects individuals and teams.

In order to understand team performance it is important to distinguish between team work and task work [32], [33], [34]. Task work refers to individual skills needed to accomplish the requirements of a specific job and team work refers to the processes and skills individuals develop to coordinate their activities. The distinction is important since the variations in outcome are, more or less, related to one of the two processes. Even both processes are extremely vital for team outcome; team work is of interest for the present purposes since task work is more related to individual skill acquisition.

The understanding of important dimensions for team work has been discussed to some extent in the research literature [35], [33], [36]. In their attempt to define the dimension that underlie effective team work (cf. [36]) researchers have taken different perspectives, such as team members personality or team work functions, as their point of departure.

The perspectives on team work dimensions often discussed is (a) the "Big Five" [37] focusing on the relation between team performance and personal traits (conscientiousness, agreeableness, openness to experience, emotional stability, and extraversion) and (b) the functional view of team performance focusing on dimensions such as decision-making, communication, and shared situational awareness (see [33] for details). Both perspectives discuss important traits or functions the teams need to have or accomplish in order to perform well.

However, even if we know that the team members need to be adaptive, share information and create a common ground [38], we also need to understand how a team can develop adaptability or become communicative in an effective way [39], [40]. In multinational coalitions, people from different countries will have to work together as a team and will be expected to perform well. These people may come from different cultures and have different values which may affect their individual behaviour and thereby the team's performance. Thus there is a need to study team performance in multinational military settings. The question is how do we capture team performance?

There are a number of factors that affect team performance. Salas et al [41] have presented a model, team effectiveness model, which describes many of the factors that have been shown to affect team performance. This model assumes that team effectiveness is a function of inputs (such as individual, team and task characteristics), and processes, such as communication and coordination. Outcomes are measured in terms of products of the collaborative process (meeting objectives or mission success), and are typically measured by the quantity and quality of the products and the efficiency and effectiveness of the process involved in producing those products. Smith-Jentsch, Johnston, and Payne, [44] present a taxonomy of the measures of interest and their relation for obtaining valid results for team performance.

In order to get a better understanding of team performance and the factors that lead to good or bad team performance there is a need to study team performance in relevant settings. However, it is difficult to find relevant platforms to study team performance, i.e. platforms where all aspects of team performance (input, process and output) can be assessed and measured systematically. Therefore a research platform called SABRE (Situational Authorable Behaviour Research Environment) was developed. Through SABRE individual and team characteristics can be collected systematically. All team communication, coordination and decision making is recorded and the team outcome is measured through Goodwill points that the team can acquire in the game if they meet the mission objectives (see method section for details).

Hypothesis 1. Significant positive correlations will be found among information sharing, situation awareness and goodwill points.

Hypothesis 2: Compared to the multi-national teams, the single nation teams will 2a) share more information (forward more tips, initiate more shared utterances, and use a greater number of map marking tools), 2b) achieve higher scores on situation awareness, and 2c) achieve more goodwill points.

Hypothesis 3: Multi-national teams will have significantly greater variability in the measures of cognitively-based cultural biases (Globesmart Commander (GC) cultural dimensions) than single nation teams.

2.2 METHOD

Following is a description of the LTAMC multicultural distributed team experiment as it was originally designed by for the ACT CD&E / HFM RTG 138 project. Deviations from the design,

though rare, did occur and are reflected in the results. Some of the material below was developed by testbed designers at BBN Technologies and documented in the SABRE NATO Experimenter's Guide. For more information about SABRE, log onto http://seriousgames.bbn.com/SABRE. This Web site contains additional information about SABRE and how it is being used for experimentation. For more information about the NATO LTAMC Experiment and SABRE, log onto http://dsl-external.bbn.com/cultural-modeling/twiki/bin/view/NATO/WebHome. The BBN material in any original and/or summarized form, is based upon work supported by the U.S. Defense Modelling and Simulation Office (DMSO) under U.S. Air Force Contract No. FA8650-04-C-6437. Any opinions, findings and conclusions or recommendations expressed in that material are those of the [BBN] author(s) and do not necessarily reflect the views of DMSO or its Contracting Agent, the Air Force Research Laboratory, Wright Patterson AFB.© 2005, 2006 BBN Technologies Corp.

2.2.1 Participants

Both culturally homogeneous and heterogeneous teams consisted of four equally ranked male officers, though different teams could have different ranks. For example, one team might consist of four Captains, whereas another team may consist of four Lieutenants. We recognize that gender can be a factor in some team interactions and perceptions of leadership. However, females played only supporting roles in the population to which findings are to be generalized. That population is the NRF DJTF. To include females in this particular study would require a significantly higher number of participants than requested or that might be available, and data from teams with female participants would, of necessity, have to be excluded from any analyses. Other requirements included an officer target age group of 18-35 with normal or corrected to normal vision, and a familiarity with computer use (e.g., mouse, keyboard). Participants must have completed undergraduate university or currently enrolled in university. Additionally, participants cannot have spent more than 6 months between the ages of 1-18 living outside of the United States and should have at least a level 3 understanding for reading, writing, and speaking English.

Eight four-person teams per nation (Bulgaria, The Netherlands, Norway, Sweden, and The United States) were required for data collection on culturally homogeneous teams for a minimum of 40 teams or 160 participants. Eight four-person teams were required for data collection on culturally heterogeneous teams for a minimum of eight teams or 32 participants. Participation in the study by any individual was limited to one session, which should occur on one day. Participant time commitment ranged from 3.5 – 5 hours, depending on the average computer skill level of the four-person team. The interactive team play of the game-based testbed meant that all four team members completed the experiment at the same time.

Participation was voluntary. All participants were given the opportunity to ask questions during the experiment. They were free to discontinue their participation at any time without penalty. All data collected was coded with subject identification numbers to protect the participant's identity. Participants' names are not linked to subject identification numbers.

2.2.2 Sample of Laboratory Layouts

Three national laboratory venues used for the LTAMC are described here (listed alphabetically by nation) to provide examples of facilities and process implemented by participating nations.

2.2.2.1 Norway

In Norway, there was employed a mobile laboratory of 5 laptop computers for the national experiments, four client machines and one server set up in a local network. We (the experimenters) controlled the server. In the experimental set-up, the 4 subjects in each group were randomly assigned to a computer. The players were sitting with their backs to each other in a large square room. Figure 1 shows the technical set-up.

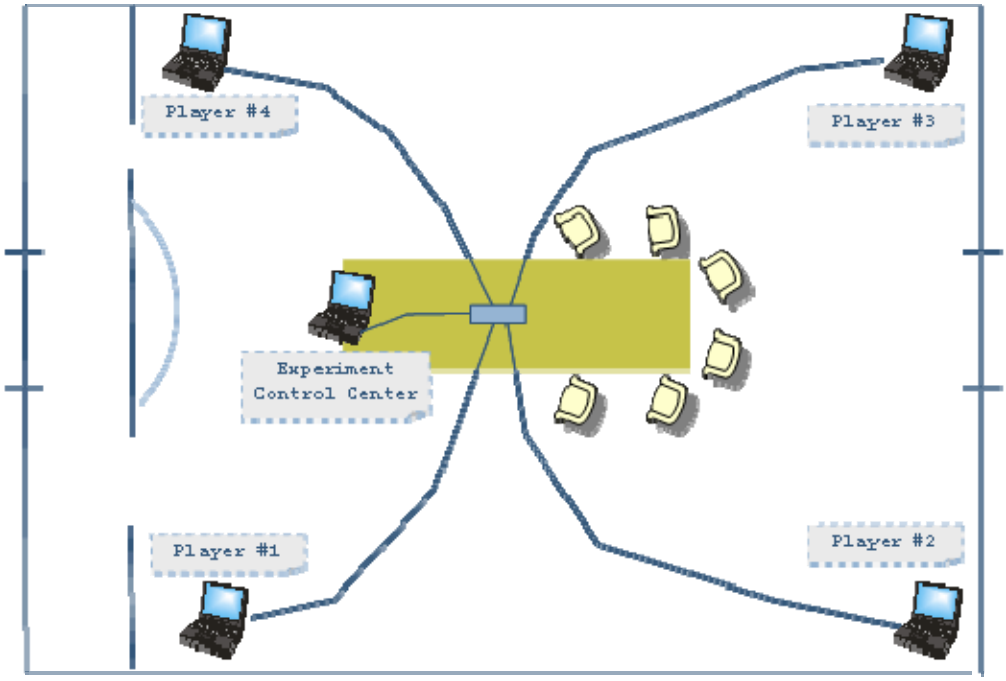

Figure 1: Technical set-up of the experiment.

There were conducted a total of 16 Norwegian national experiments; the first series of 8 experiments were conducted in a class-room at a higher Norwegian military educational institution (master level) and the second series of 8 experiments were conducted in a class-room at a lower Norwegian military educational institution (bachelor level). The first series of national experiments were conducted on 8 days between 24.04 and 15.05 2006 and the second between 12.10 and 10.11 2006.

The international experiments were conducted on a regular stationary computer located in an office at the Norwegian Defense Research Establishment (FFI). Norway participated in 6 of these, executed between 06.03 and 28.03.07. Subjects were recruited from a lower Norwegian military educational institution (bachelor level).

2.2.2.2 Sweden

Swedish data was collected at five sites in Sweden to facilitate volunteer participation:

Army Combat School Kvarn, Linköping. The school instructs, develops and trains individuals, units and systems in ground combat. The school has an advanced training facility to train Military Operations in Urban Terrain. Researchers provided NeverWinter™ Nights and SABRE, using computers, network equipment, monitors, keyboards and headsets provided by the school. Experimental sessions were conducted in the school's combat technique laboratory over the course of two days.

Army Combat School, Skövde. The school instructs, develops and trains individuals, units and systems in ground combat. The school has unique education systems where simulators, adapted technology increase realism and training effectiveness. The school has an advanced Crew Training Facility for tank personnel and a Command Training Facility for commanders and their staffs. For the experiment researchers provided laptop

computers, network equipment and headsets, using monitors and keyboards provided by the school. Experimental sessions were conducted in a classroom over the course of two days.

Swedish Armed Forces Development Centre, Enköping. At this facility the Swedish Armed Forces conduct experiments, studies and tests. It is an important facility for CD&E and MNE related studies. It is a large facility where different technologies and methodologies are implemented and tested. For the experiment researchers provided Never Winter Nights and SABRE, using computers, network equipment, monitors, keyboards and headsets provided by the facility. Experimental sessions were conducted during one day.

Norrbotten Wing, Luleå. This regiment creates Task Forces; Squadrons, Battalions, fighter control battalions and signal and communication units. The regiment also takes part in surveillance of Swedish air territory. For the experiment researchers provided laptop computers, network equipment and headsets, using monitors and keyboards provided by the regiment. Experimental sessions were conducted during one day.

Swedish Defence Research Agency (FOI), Stockholm. FOI, the Division of Command and Control Systems, was the sponsoring organization in Sweden for the data collection effort. Research in this division of FOI covers areas such as Command and Control Systems, electronic warfare, IT security and communications. The aim of this division is to design systems which make full use of human cognitive potential. The division of Command and Control Systems is located in Linköping. Due to practical reasons the data was collected at FOI in Stockholm. FOI in both Stockholm and Linköping have laboratories for advanced simulations and data collection. The experiment was set up with equipment from FOI and data was collected.

2.2.2.3 The United States

U.S. data were collected at the U.S. Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Interface Division, Cognitive Systems Branch Culture and Cognition Laboratory (CCL), Wright-Patterson Air Force Base, Dayton Ohio USA. The lab contained five desktop computers: four of which were client computers located in one room and one server computer located in an adjoining room. For data collection on culturally homogeneous teams, the effect of distributed teamwork was achieved by placing physical barriers (portable walls) between participants. This arrangement meant that participants could not see each other nor see each other's computer screens. To further create the feeling of being on a distributed team, all participants wore headsets during the experiment session. With culturally heterogeneous teams, participants were truly distributed with only one person from any given nation sitting at the client computer. Assisting the Principle Investigator, were two lab assistants currently enrolled in the Psychology graduate program at Wright State University and two computer technicians.

2.2.3 SABRE Game-Based Testbed

Participants played a non-violent interactive computer video game modified by BBN Technologies from Bioware's Neverwinter Nights using the game's provided toolset [43]. This game-based testbed, titled "Situation Authorable Behaviour Research Environment (SABRE)," required participants to make responses using the keyboard and mouse while working in a team composed of four people. The primary goal of the specific SABRE game modification used for the NATO experiment was to search for hidden supplies of weapons in a virtual city and, in the process, earn or lose the goodwill of the non-player characters (NPC) (i.e., townspeople). Participants communicated by typing messages in a chat box and were not in visual line-of-sight with their teammates. They wore headsets to prevent distraction from the experimental task.

2.2.4 Phases of Game-Play

The NATO experiment scenario required exactly four participants, with one participant randomly selected to be the team leader at the beginning of each experimental session. The team was given the mission of searching for hidden weapons caches inside and outside of buildings in a virtual city. As a result of the searches, a team Goodwill score was generated. The Commander's Intent, delivered by a non-player character in the game, was for the team to maximize their Goodwill score as they searched the virtual city for weapons caches. Each team began game play with the same number of Goodwill points.

After completing four computer-administered, pre-game surveys, participants were signed on to the *Training Phase* of the game by the experimenter. During the training (familiarization) phase of the experiment, participants go through a series of virtual areas (e.g., rooms, hallways), each designed to teach participants individually about different aspects of the game. After showing proficiency on individual tasks, participants are brought together by the game and taught team skills. The training phase is not time constrained. Typically, participants required 1 to 2 hours to complete the training phase.

After completing the team skills section of the Training Phase, participants were signed on to the *Planning Phase* by the experimenter. The Planning Phase occurred at the start of the game, before the team executed the mission. The Planning Phase began the actual experiment. In this phase, participants were first asked to discuss, using text chat, the distribution of five responsibilities among the team. Each responsibility could have been taken by no one person, one person, or more-than-one person. The responsibilities consisted of: Coordinating the search plan; Monitoring the Goodwill score; Monitoring the time elapsed and remaining; Tracking the number of weapons caches located; and Managing information. Next, the team is required to distribute equipment that will aid in finding weapons caches. The team equipment included: two high-fidelity sensors, one lock pick, one Goodwill monitor, one time monitor, one re-planning trigger, and one explosives/IED detector. Also occurring during the Planning Phase, each participant was provided with a map marking tool and a journal management tool. Participants were each given two distinct tips about where to search for caches. At this point, and at any point in the game, team members could share or assign their tips to their teammates by using the tip management tool. Then, within the game, they were shown a map of the town and asked to formulate a search plan and a communications plan. A "spokesperson" (not necessarily the randomly selected leader) was selected by the team to summarize each of the two plans for the Commander. The experiment will not progress until both plans are documented. Finally, each participant's journal is automatically updated with the group plans, their own responsibilities, and the overall equipment distribution. The planning phase is not time constrained. Typically, participants will require approximately 30 minutes for completion.

Upon completion of the Planning Phase, participants immediately began the *Execution Phase* of the mission, where the team enters the virtual town and begins the task of locating and acquiring caches of weapons. The exchange of information among team members could be either structured or unstructured. Team members could communicate with each other through the structured formats of their map-marking tool and journal-management tool. They could also communicate by typing text messages (unstructured). When participants used the chat *Talk* mode, they sent a broadcast message to all nearby teammates. When participants used the chat *Tell* mode, they sent targeted, long-distance messages to a specific teammate.

Three times during game-play, at 20-minute intervals, participants were presented within-game situation awareness probes. All participants had to independently enter their responses to the probes before game play could continue. The Execution Phase of the experiment was timed, giving participants one hour to complete the task at which time the game stopped automatically.

2.2.5 Dependent Variables

2.2.5.1 Information Sharing

The amount of information sharing for a team was measured by the sum of three sub-measures: the number of typed utterances initiated, total number of information tips forwarded, and total number of map marker uses. Map markers were used by team members to indicate that a building or area had been searched.

2.2.5.2 Situation Awareness

The team's situation awareness score was the aggregate scores over three situation-awareness probes delivered to each participant during the experiment. There were a total of six automatically scored questions on each situation awareness probe. Because there were four team members, and the probe occurred three times during the experiment, the maximum team score was 72 correct answers.

2.2.5.3 Team Performance

The team's performance was measured by difference in Goodwill points that they started out with and their final Goodwill score (i.e., the Delta). Outdoor caches were worth 100 Goodwill points and indoor caches were worth 300 Goodwill points. The main costs to Goodwill points were associated with entering private buildings to search for caches. Because the team was provided with the tools to screen locations for weapons prior to taking action, it cost 50 points if a private building was entered and no weapon cache was found (whether there was none to be found, or they failed to find it). Participants were penalized 100 points if they opened a container that did not contain a cache. They were also penalized 250 points if they set off one of the two IEDs in the game (by opening a rigged crate). Only interior (i.e., inside a building) crates had IEDs. If a participant opens a rigged crate, a small yellow light appeared around the crate then disappeared from the screen after a few seconds. The player character, or avatar, was not physically harmed by the explosion and does not even jump when the soundless "explosion" occurred. The team was provided with information, equipment, monitors, and tools to assist with searches and help maintain team situation awareness. There were no individual Goodwill scores. All participants were informed each time any individual team member gained or lost Goodwill points for the team.

The maximum number of weapons that the team could find in the town was 20. That included 4 outdoor, 12 indoor, and 4 indoor that existed only for a limited time before "disappearing." (Tips were provided in-game regarding the caches that had a limited life-span.) This scoring could gain the team 300 Goodwill points for each indoor weapons cache, and 100 for each outdoor cache, for a maximum gain of 5800 Goodwill points. The team could also gain a up to 530 Goodwill points if they completed a variety of sub-quests, thus earning extra goodwill from the non-player character townspeople. Although the theoretical maximum performance for a team would be to gain 6330 Goodwill points, it is unlikely that a team could manage to do this. Similarly, because there are approximately 40 houses, 10 empty crates, and 2 trapped crates, a team could theoretically lose 3500 Goodwill points. It is unlikely that a team would do that either.

2.2.6 Survey Instruments

Several in-game surveys were administered. They include a background and a debrief questionnaire developed by the research team, the NEO-FFI Personality Inventory [44], [45]; Hofstede Value Survey Model 1994 Culture Survey [46]; Organizational and Interpersonal Factors questionnaire developed by project team member Ms. Anne Lisa Bjornstad (Norwegian Defence Research Establishment, FFI); and the GlobeSmart® Commander Culture Survey (Copyright: [47], [48]).

2.3 PRIMARY RESULTS

2.3.1 Hypothesis 1. Significant positive correlations will be found among information sharing, situation awareness and goodwill points.

Table 1 presents the correlation matrix for the dependent variables: Information (Info)-Sharing: Briefing Phase and Residential Search Phase, three self-reports of situation awareness (SA1, SA2, and SA3), and the Net Goodwill Balance. Hypothesis 1 was partially supported. As expected, team members' information sharing in the briefing phase was strongly related to information sharing in the residential search phase, and had a weak, but significant relationship later in the search phase with situation awareness (SA 3) and goodwill points. Information sharing in the residential search phase had a weak significant relationship with situation awareness (SA 1, SA 2, and SA 3) throughout that phase. In contrast, no relationship was found between goodwill points and information sharing in the residential search phase or with goodwill points and situation awareness.

Table 1: Pearson correlation matrix for the dependent variables (Information-Sharing: Briefing and Residential Phases, SA, and Net Goodwill Balance).

	Info Sharing Briefing	Info Sharing Residential	SA1	SA2	SA 3
Info Sharing: Residential	.576**				
SA 1	.115.	.178**			
SA 2	.123.	.179**	.519**		
SA 3	.130*	.136*	.433**	.448**	
Net Goodwill Balance	.129*	.056	.079	.111	.066

***N* = 224, * *p* < .05, * *p* < .01**

2.3.2 Hypothesis 2: Compared to the multi-national teams, the single nation teams will 2a) share more information (forward more tips, initiate more shared utterances, and use a greater number of map marking tools), 2b) achieve higher scores on situation awareness, and 2c) achieve more goodwill points.

Table 2 shows the mean and standard deviation for each dependent variable by national and multi-national teams: Information (Info)-Sharing: Briefing and Residential Phases, three reports of situation awareness (SA), and the Net Goodwill Balance. The total number (N) of participants is listed next to each nation. For all of the analyses conducted in this section, the partial eta squared (η^2) statistic is included for significant effects. All post hoc analyses used the Tukey Honestly Significant Difference for Unequal N statistic.

2.3.2.1 Information-Sharing: Briefing Phase

The Information-Sharing: Briefing Phase data was analyzed within a 1-way ANOVA with nations as the independent variable. The analysis revealed hypothesis 2a was partially supported. A significant effect of nations was found, $F(5, 218) = 15.07$, $Mse = 533.74$, $p < .001$, and $\eta^2 = .26$. Figure 2 shows the national and multi-national total scores for information shared during the SABRE Task Briefing Phase. The post hoc analysis showed that, statistically, the Norwegian and Bulgarian teams did not differ in the amount of information shared during the briefing phase within their teams and demonstrated the least amount of sharing. In contrast, the Dutch and American teams had the highest scores for information sharing in the briefing phase, followed by the Swedish and multi-national teams, who did not differ statistically in the amount of information shared within their teams.

Table 2: Mean and standard deviation for each dependent variable by national and multi-nation al teams: Information-sharing: briefing and residential search phases, three reports of Situation Awareness, and Net Goodwill Balance.

Nations (*N*)	Info-Sharing: Briefing Phase (std err)	Info-Sharing: Residential Phase (std err)	SA Report 1 (std err)	SA Report 2 (std err)	SA Report 3 (std err)	Net Goodwill Balance (std err)
Bulgaria (32)	18.06 (2.13)	15.63 (1.69)	4.50 (.27)	4.53 (.23)	4.37 (.19)	186.56 (78.78)
Netherlands (32)	59.97 (5.17)	62.75 (5.16)	4.63 (.23)	4.53 (.23)	4.18 (.19)	250.31 (80.89)
Norway (64)	33.73 (2.44)	49.36 (3.49)	4.66 (.12)	4.56 (.12)	4.17 (.14)	142.50 (31.40)
Sweden (36)	42.22 (3.05)	56.16 (3.09)	4.67 (.22)	4.69 (.24)	4.44 (.21)	190.00 (61.75)
U.S. (28)	56.93 (6.35)	55.64 (6.03)	3.75 (.24)	4.07 (.29)	4.14 (.23)	308.93 (113.53)
Multi-National (32)	48.19 4.41	43.03 (3.68)	4.46 (.23)	4.34 (.21)	4.34 (.23)	265.31 (78.31)

2.3.2.2 Information-Sharing: Residential Search Phase

The Information-Sharing: Residential Search Phase data was analyzed within a 1-way ANOVA with nations as the independent variable. The analysis revealed hypothesis 2a was partially supported. A significant effect of nations was found, $F(5,218) = 15.07$, $Mse = 602.39$, $p < .01$, and $\eta^2 = .26$. Figure 2 shows the National and Multi-national total scores for information shared during the SABRE Task Residential Search Phase. The post hoc analysis showed that during the search phase, the Bulgarian teams demonstrated the lowest amount of information sharing with team members, whereas the Dutch, Swedish, and American teams showed the greatest amount of information sharing. The Norwegian teams had the second lowest amount of information

sharing in the search phase and the multi-national teams did not differ significantly from the Norwegian, Swedish, and U.S. teams in their amount of information sharing. Since the experiment was conducted in English and involved a complex computer game, we would expect these factors to affect the pattern of results. A covariate analysis, which also included age since age negatively correlates with game experience, indicated that these three factors together accounted for 40% of the variance in performance scores (Warren, 2008). When the effects of these confounding factors were removed, all nations had comparable median performance and the mixed-nation teams as a whole had the best performance.

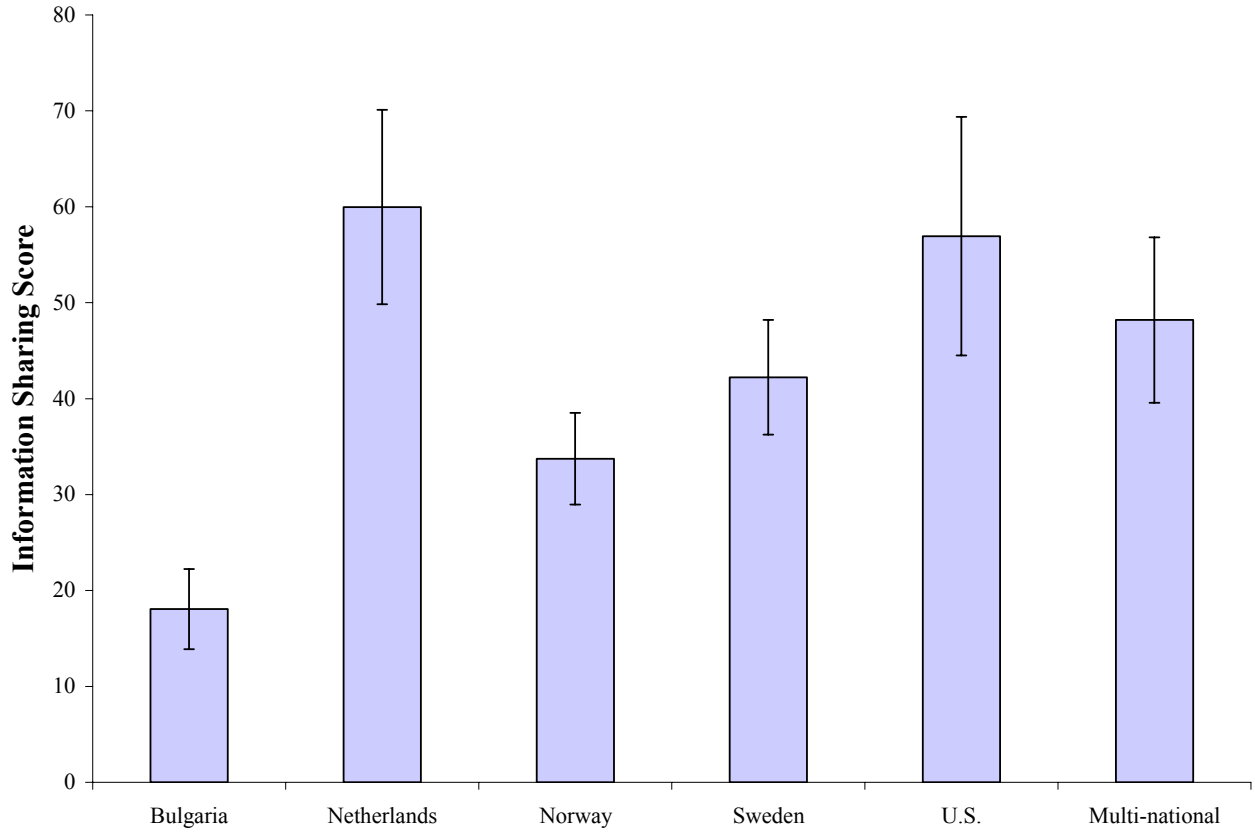

Figure 2: National and multi-national total scores for information shared during the SABRE Task Briefing Phase.

2.3.2.3 Situation Awareness

Hypothesis 2b was not supported. The SA data was analyzed within a 6 (nations) x 3 (SA reports) repeated measures analysis of variance (ANOVA) with repeated measures on the last variable. The analysis revealed null effects of nations ($F(1, 218) = 1.38$, *ns*) and SA (Wilk's $\lambda = .98$, $F(2, 217) = 2.44$, *ns*), as well as a non-significant interaction between nations and SA (Wilk's $\lambda = .95$, $F(10, 434) = 1.09$, *ns*).

2.3.2.4 Net Goodwill Balance

Hypothesis 2c was not supported. The Net Goodwill Balance data was analyzed within a 1-way ANOVA with nations as the independent variable. Similar to the SA analysis, there was a null effect of nations on the Goodwill scores ($F < 1$).

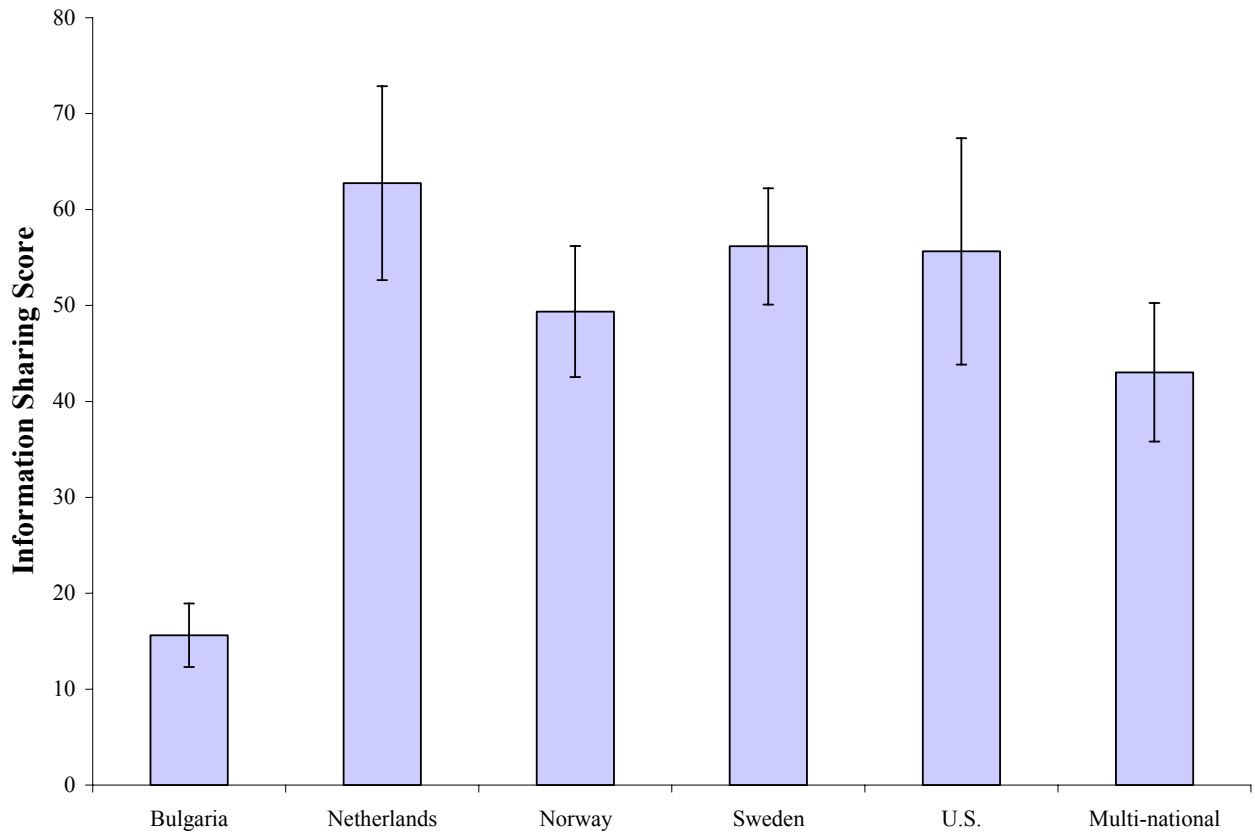

Figure 3: National and Multi-national total scores for information shared during the SABRE Task Residential Search Phase.

2.3.3 Hypothesis 3: Multi-national teams will have significantly greater variability in the measures of cognitively-based cultural biases (GlobeSmart® Commander cultural dimensions) than single nation teams.

An inter-rater agreement statistic was calculated (i.e., the average deviation index, or AD) to examine the degree of agreement among team members within the same team with respect to their responses to each GlobeSmart® Commander cultural dimension. We computed the AD index and its associated test of statistical significance for each team's responses to each dimension. We then calculated, for each team, the proportion of items on each dimension for which the team members had non-chance (i.e., statistically significant) levels of agreement. Hypothesis 3 was supported. Across all dimensions, the multi-nation teams consistently had the lowest level of agreement on each dimension compared to the single nation teams. Single nation teams were uniformly agreeable on all but the egalitarian dimension, having significant levels of agreement (i.e., average deviation statistic with a $p < .05$) on approximately 80% of the items for any one dimension. For example, on the ST-LT dimension, Norwegians were most close in their agreement on the scale items (83.50%) while the multi-nation teams had the lowest agreement (54.25%). Only the egalitarian/status dimension resulted in less agreement within individual nations. For the NEO-FFI dimensions, internal consistency reliabilities as indicated by Cronbach's Alpha ranged from .66 - .83, which is roughly consistent with previous research (Costa & McCrae, 1990).

The effect of these findings may be reflected in the poor internal consistency reliability estimates calculated for each of the GlobeSmart® Commander dimensions. Cronbach's Alpha was calculated using the entire group of participants in the experiment (N=223). The estimates were found to be lower than is considered reasonable for self-report individual difference measures [49].Internal consistency reliabilities for the original scales ranged from .09 - .53. After removing problematic items with the lowest corrected item-total correlations, reliabilities ranged from .35 -.63. Only the GlobeSmart® Commander Short Term-Long Term (ST-LT) culture dimension had the highest corrected Cronbach's Alpha of .63. The low internal consistency of each scale may be due to mixing the single nation and multi-nation data together. Further item analyses on the GlobeSmart® Commander culture scales are available upon request.

2.4 SECONDARY RESULTS

2.4.1 Descriptive Statistics for All Study Variables

All of the individual difference measures were reasonably distributed with some skewness and some kurtosis. On the performance variables, most of the interval level data was negatively skewed (i.e., a high frequency of low numbers around zero). For the 3 SA reports, the distribution was slightly positively skewed. Box plots and frequency distributions are available upon request.

Significant correlations among each personality dimension, and each GlobeSmart® Commander dimension, and the SABRE performance outcomes are described below. Correlations reported are significant at the $p < .05$ level. The full matrix is available upon request.

Neuroticism:

- + Independent: individuals who are higher on neuroticism also tended to be more independent.
- + Risk: individuals who are higher on neuroticism also tended to be more willing to take risks.
- - Direct: individuals who are higher on neuroticism also tended to be less direct.
- - Short: individuals who are higher on neuroticism also tended to think short term.

- - Information sharing – briefing: individuals who are higher on neuroticism also tended to share less information.

Extraversion:

- - Risk: individuals higher on extraversion tended to be less likely to take risks.
- + Direct: individuals higher on extraversion tended to have a more direct communication style.
- + Forwarded tip count: individuals higher on extraversion tended to forward more tips to others.

Openness:

- - Risk: individuals higher on openness tended to be less likely to take risks.
- + Short: individuals higher on openness tended to think more short-term.
- + Information sharing: individuals higher on openness tended to share more information with others.

Agreeableness:

- - Independent: more agreeable people also tended to be less independent.
- + Short: more agreeable people also tended to think more short-term.
- + Age: more agreeable people also tended to be older.

Conscientiousness:

- - Risk: more conscientious people also tended to be less likely to take risks.
- + Short: more conscientious people also tended to think more short term.

Independent-Interdependent: No significant correlations observed.

Egalitarianism-Status:

- + Exterior cache finds: more egalitarian people also tended to find more exterior caches.
- - Chat count: more egalitarian people also tended to chat less with other team members.
- + Map marking tool count: more egalitarian people also tended to use the map marking tool more.

Risk-restraint: no significant correlations observed, though a marginal correlation with the SA score report 1 (p <.10) was noted – this effect disappeared by SA report #3.

Task-Relationship:

- + Exterior cache finds: more task oriented people also tended to find more exterior caches.
- + Interior cache finds: more task oriented people also tended to find more interior caches.
- + Net good-will balance: more task oriented people also tended to have higher good will balances.

Direct-Indirect:

- + Forwarded tip count: more direct people also tended to forward more tips to others.
- - Map marking tool count: more direct people were less likely to use the map- marking tool.

- + Utterance count: more direct people also tended to have more utterances.

Short-term-Long-term:

- + Information sharing – briefing phase: those with more short-term orientation were also more likely to share information.

2.4.2 Team Homogeneity Analyses

We calculated inter-rater agreement statistics (i.e., the average deviation index, or AD) to examine the degree of correspondence among team members within the same team with respect to their responses on individual GlobeSmart® Commander items. We computed the AD index and its associated test of statistical significant for each team's responses to each item on the GlobeSmart® Commander dimensions. We then calculated, for each team, the proportion of items on each dimension for which the team members had non-chance (i.e., statistically significant) levels of agreement.

For all dimensions except for the egalitarianism dimension, teams were uniformly homogenous on the GlobeSmart® Commander dimensions, having significant levels of agreement (i.e., average deviation statistic with a $p < .05$) on approximately 80% of the items for any one dimension.

For the egalitarianism dimension, there were differences among teams with respect to their homogeneity. Norwegian teams were most homogenous (i.e., 86% of items had significant agreement), and mixed teams (i.e., 54% of items had significant agreement) were least homogenous.

2.4.3 Comparison of Observed GlobeSmart® Commander Scores to National Norms Reported on GlobeSmart® Website

We further calculated team means for each item on each dimension, as well as 95% confidence intervals around these means. We then examined the norms for each nationality reported on the GlobeSmart® website, and used the dimension means as benchmarks against which to compare our observed team means. For each item, if the team's confidence interval included the benchmark value for that team's nationality, the item was considered a 'hit', if the benchmark value was not included in the confidence interval it was considered a 'miss'. We then calculated, for each team, the proportion of items for each dimension that were hits.

Teams differed consistently on their proportion of hits as a function of their nationality.

- Independence: Americans were highest and Bulgarians were lowest.
- Egalitarianism: Swedes were the highest and Bulgarians were the lowest.
- Risk: Americans were the highest and Swedes were lowest.
- Task: Americans were highest and Bulgarians were lowest.
- Direct: Dutch were highest and Bulgarians were lowest.
- Short term: Americans were highest and Swedes were lowest.

2.4.4 Item Response Theory Analyses

The GlobeSmart® Commander items, like many self-report measures of individual differences (i.e., designed to assess values, attitudes, or personality), can be considered examples of ideal-point, or Thurstonian items. With ideal-point items, the probability of a positive response to the item depends on the difference between the individual's trait level and the position of the item on the trait continuum (traditionally considered the

"difficulty" level of the item): $\theta - \delta$, and is maximized when the individual and the item are on the same point on the trait continuum. Thus, the probability of a positive response increases non-monotonically (i.e., like the normal curve) with the respondent's actual trait level. Often, Thurstonian items consist of statements with which an individual can disagree for two reasons: 1) the individual's trait level is below that targeted by the item, or 2) the individual's trait level is above that targeted by the item. For example, consider the item "I tend to take on extra work tasks if I know my supervisor will hear about it" designed to measure work motivation. Individuals too low on the trait would disagree because they never take on extra work tasks, whereas those too high on the trait would disagree because they take on extra tasks even if their supervisor might not hear about it. Because the item targets a moderate level of the trait, individuals with moderate levels of the trait would have the highest probability of response.

Ideal-point items can be contrasted with cognitive ability items, in which the probability of a correct answer increases monotonically with higher levels of the underlying ability. For example, someone with high math ability will have a high probability of a correct response to a moderately difficult math item no matter how much higher their trait level vis-à-vis the trait level demanded by the item. In other words, one can never have "too much" of the trait.

Responses to ideal-point items are best modelled using unfolding item response theory (IRT) models. IRT models are a family of psychometric models that allow one to estimate the probability of a positive response to an item as a function of the where the person and the item are located on the latent trait continuum. Unfolding IRT models take into account the marked folding in the item characteristic curves that occurs for moderate items (i.e., the respondent can disagree from above or below). It is important to note that unfolding models will yield similar results as cumulative IRT models (i.e., IRT models used traditionally for ability-type items) for extreme items (e.g., "Finding time to organize your belongings is your top priority).

The primary benefit of ideal-point items and unfolding IRT models is that they ensure increased measurement precision across all levels of the trait continuum. It is difficult to estimate trait levels using cumulative, monotonic IRT models for individuals who are in the middle of the trait continuum. Unfolding IRT models alleviate this problem by capturing the folding that appears in the extremes of the trait continuum, in which case cumulative models would lead to underestimation of trait levels for individuals with the most extreme values of the latent trait. Thus, unfolding models ensure the benefits associated with IRT (e.g., person-invariant item parameters, item-invariant trait estimates, high measurement precision) but provide a better fit to the data for neutral/moderate items, ensuring accurate trait estimation across levels of the trait continuum.

We used the Generalized Graded Unfolding Model (i.e., "GGUM"; Roberts, 2002) to estimate item parameters for each of the GlobeSmart® Commander dimensions and to estimate individuals' latent trait levels along each of the dimensions. The original response scale, with eight graded response options, was transformed to a new scale with four response options. Adjacent categories were combined (i.e., original 1 and 2 = 1, 3 and 4 = 2, 5 and 6 = 3, 7 and 8 = 4). The fully specified GGUM was found to fit the data adequately for each dimension.

We estimated the following item parameters for each item:

- Item difficulty – where the item stand on the latent trait continuum.
- Discrimination – the slope of the item characteristic curves and the ability of each item to discriminate between individuals high and low on the respective dimension.

- Response category threshold position – the point on the latent trait continuum at which the probability of responding in the next highest category (e.g., transitioning from agree to strongly agree) becomes higher than the probability of responding in the current category.
- Response category threshold distance – the distance between response category thresholds.

Models estimated for each dimension were found to fit the data adequately. Item fit and person fit statistics were found to be adequate for each scale. Estimated trait levels take into account item parameters and vary along a - 4 to + 4 scale, as is the standard for IRT analyses. Item characteristic curves are available upon request.

2.4.5 Regression Analyses

We performed a set of multiple regression analyses to examine 1) the direct influence of the set of cultural dimensions on multiple aspects of performance (using the IRT trait estimates for each GlobeSmart® Commander dimension), and the incremental validity of 2) team nationality, and 3) personality, as measured by the NEO-FFI. These analyses address the extent to which team nationality can be used as a proxy for team culture as measured by the GlobeSmart® Commander. By independently assessing nationality, culture, and personality, we can address their independent effects and potential causal orderings among these variables. Results are described by SABRE performance variable.

2.4.5.1 Cache Finds

- Interior cache finds: the set of GlobeSmart® Commander dimensions as a whole accounts for approximately 10% of the variance in interior cache finds. Adding team nationality into the model does not account for significantly incremental variance in interior cache finds.
- Exterior cache finds: the same pattern was observed as with internal cache finds, except that the variables accounted for 7% of the variance.
- Although there were no significant differences on exterior cache finds, American and Mixed teams had the highest mean numbers of cache finds.
- Personality accounts for an additional 4% of the variance above and beyond the effects of the GlobeSmart® Commander dimensions for interior but not exterior cache finds. Agreeableness and Conscientiousness appear to be the best predictors.

2.4.5.2 Chat Count

- Team nationality predicts chat count, but it the effects of nationality are not transmitted through national differences in culture.
- Swedish and Dutch have highest number of chat counts.
- Bulgarian and Mixed groups have the lowest number of chat counts.
- Personality does not predict above and beyond the effects of team nationality.

2.4.5.3 Tip Count

- The set of cultural dimensions as a whole did not predict tip count, but the direct-indirect dimension was a significant predictor of tip count. Those who are more direct tend to forward more tips.
- The incremental validity of team nationality is marginally significant at $p = .07$.
- Swedish and American teams had the highest number of Tip counts.

- Bulgarian and Norwegian have the lowest number of Tip counts.
- Personality accounts for an additional 5% of the variance above and beyond team nationality. In particular, extraversion and conscientiousness are the best predictors. Higher extraversion and lower conscientiousness are associated with higher numbers of tip counts.

2.4.5.4 Information Sharing

- For the briefing phase, neither the set of cultural dimensions nor team nationality predicted information sharing.
- In the search phase, the cultural dimensions did not predict, but nationality did predict information sharing. Dutch, Swedish, and Americans had the highest scores and Bulgarian and Mixed had the lowest.
- Personality did not add significant prediction above and beyond nationality

2.4.5.5 Map Tool Markings

- The set of cultural dimensions accounts for a significant amount (10%) of the variance in scores for map tool markings.
- The addition of team nationality into the model does not increase prediction.
- Even though team nationality did not offer incremental prediction above and beyond the effects of the set of cultural dimensions, there was a significant difference in map tool markings by team nationality. Dutch and Bulgarian were the highest and Norwegian and Swedish were the lowest.
- Personality did not predict map tool markings above and beyond the effects of culture.

2.4.5.6 Utterances

- Among the cultural dimensions, direct-indirect was the only significant predictor of the number of utterances, and team nationality predicted utterances above and beyond the effects of the direct-indirect dimension. This implies a partial mediation effect in which the influence of team nationality on number of utterances is partially transmitted through scores on the direct-indirect dimension.
- Test of the partial mediation model: Differences in directness-indirectness associated with team nationality affected the number of utterances.
    - Teams with high scores on the dimension (i.e., most direct), Swedish, had highest numbers of utterance counts.
    - Teams with lowest scores on the dimension, Bulgarian, had lowest numbers of utterance counts.
    - Personality does not predict above and beyond the effects of culture.

2.4.5.7 Goodwill Points

- For net-goodwill balance, the task-relationship dimension was the only one that emerged as a significant predictor.
- No reliable differences were associated with team nationality.
- Personality did not predict above and beyond the effects of culture.

2.4.5.8 Situational Awareness

- No reliable predictors of situational awareness were observed for any of the three reports.

- Lower levels of agreeableness associated with higher levels of SA for report 1 only.

2.5 DISCUSSION

Each of the hypotheses we addressed in this chapter were supported, but with some important exceptions. Weak relationships were found among measures of team process and outcome. More information sharing in the briefing phase weakly predicted greater situation awareness and higher goodwill points. In contrast more information sharing in the residential phase was only weakly related to greater situation awareness and situation awareness was not related to goodwill points.

The finding that only information sharing varied across nationality suggests that a more complex relationship may be moderating or mediating relationships among the dependent variables. The GlobeSmart® Commander cultural dimensions may be influencing relationships, as the results suggest that perceptions about the dimensions are stronger within single nation teams than in multi-national teams. Further analyses should be conducted to determine whether more complex relationships may exist.

Further, subsequent analysis has revealed a strong mediating affect of language competency and age on the impact of culture on teamwork. Specifically, when age and language are used as covariates in the data analyses, performance differences between nations disappear.

2.6 CONCLUSIONS

This experiment represents a highly significant accomplishment that has never been done before. It has established a baseline comparing team performance data from 56 teams representing five separate nations and multi-national teams comprised of members from these same nations. The findings bear out the general principle that information sharing in teams is related to situation awareness, and that the process of information sharing predicts team performance outcomes. Nationality influences information sharing, but not situation awareness or team outcomes, and it is not yet clear what the nation specific factor might be. Further exploration of cognitive cultural biases using the GlobeSmart® Commander culture dimensions measure is needed to understand the effect of nationality on the psychometric properties of the measure. Such analyses are needed before it can be used to identify mediating or moderating effects of culture on team processes and outcomes.