9-9-2005

Does Agreeableness Help a Team Perform a Problem Solving Task?

Frederick R. B Stilson
University of South Florida

Does Agreeableness Help a Team Perform a Problem Solving Task?

by

Frederick R. B. Stilson

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Michael D. Coovert, Ph.D.
Carnot Nelson, Ph.D.
Marcia A. Finkelstein, Ph.D.

Date of Approval:
September 9, 2005

Keywords: personality, big five, computer, simulation, critical thinking

Dedication

This Masters Thesis is dedicated to my family and friends, especially my parents, Fred and Sharon Stilson, who always told me that I could accomplish anything to which I put my mind.

Acknowledgments

I would like to thank the following people, without whom this thesis would have never been possible:

Dr. Michael Coovert
Dr. Carnot Nelson
Dr. Marcia Finkelstein
Tim Willis
Matt Prewett
Ashley Gray
Rebecca Klein
Katie Bush
and
Aptima

Table of Contents

List of Tables

List of Figures

Does Agreeableness Help a Team Perform a Problem Solving Task?

Frederick R. B. Stilson

ABSTRACT

The relationship between mean team *Agreeableness* and team performance has not been shown definitively. The present study was performed looking at archival data from a study that assessed team performance from 62 two person teams using the DDD and involving two types of training and two types of information probes during the computer task. In addition, each of the participants took a personality test based on the IPIP with an emphasis on *Agreeableness* and its 6 facets. Using HLM analysis, it was determined that *Agreeableness* does not have a significant effect on team performance for a problem solving tasks ($\Delta\chi^2$=2.04, p=n.s.), however it did significantly effect how an individual performed ($\Delta\chi^2$=18.06, p=.001) on the problem solving task. Intelligence had a significant effect on team performance ($\Delta\chi^2$=569.08, p=.001) and this may have washed out any personality effects. In addition, a linear regression indicated than none of the six facets of *Agreeableness* had a significant effect on team performance on a problem solving task.

In recent years, psychologists have used personality measurements like the International Personality Item Pool (IPIP) (Goldberg, 1999) to predict many different dependent variables. Psychologists in several different fields have studied personality using tools like the IPIP and then observing how personality relates to everything from psychosis to job performance. Beginning in the 1980's a structured approach to defining personality began to emerge: the Five Factor Model (FFM). This new formulation and the term FFM were coined by McCrae and Costa (1985). The FFM includes the personality dimensions of Openness to Experience, Conscientiousness, Extraversion, *Agreeableness*, and Neuroticism, commonly referred to as the acronym OCEAN. Conscientiousness, Extraversion and Neuroticism have been studied the most with fairly consistent and stable results emerging. In contrast to these three personality factors, *Agreeableness* and Openness to Experience are often overlooked areas of the FFM. Specifically, *Agreeableness*, its facets, and team performance on a computer simulated task have been overlooked. To define team personality in the current study, methods utilized in past studies, such as average scores (Neuman and Wright, 1999) and individual means (e.g. Heslin, 1964; Williams & Sternberg, 1988; Barrick, Stewart, Neubert, and Mount, 1998) were used.

This study paired a personality test with a computer simulated task in order to help fill that knowledge gap in literature. A computer simulated task was chosen for this study because it allowed the participants to be presented with a novel situation. If one were to use a task one might encounter in the business world, a possible confound of having participants with specific training in that area of business may have arisen. In addition to personality and computer simulations, the literature review for the current study will specifically cover the following topics: the Big Five Personality factors, findings involving team learning and personality and specifically how team learning and team performance are related, teams and backing up behavior, why a computer simulated task is relevant, the relationship between team performance and team learning, and teams and the shared mental model. In addition, an overview of the ambiguous results involving Openness to Experience, *Agreeableness* and team performance is outlined and finally, the current study, one that examined individual and average team *Agreeableness* and its various facets and how they affected team performance on a computer simulated task, is discussed.

The Big Five Personality Factors

The concept of the Big Five Personality Factors was originally developed by Tupes and Christal (1961) and Norman (1963). This concept has been subsequently confirmed by Goldberg (1999) and McCrae and Costa (1985). The Big Five Personality concepts consist of Openness to Experience, Conscientious, Extraversion, *Agreeableness,* and Neuroticism and together the five are commonly referred to as OCEAN.

Openness to Experience involves being imaginative, curious, having broad interest and possibly going about life in an untraditional manner. Conscientious refers to one who is organized, punctual, ambitious and persevering. Extraversion is a trait marked by being sociable, talkative, fun-loving, and optimistic, which is in contrast to the fourth trait of Neuroticism. One who is classified as Neurotic is often worrying, nervous, and insecure. Finally, *Agreeableness* is shown in someone who is soft- hearted, helpful, forgiving, and possibly gullible (Costa, Busch, Zonderman, and McCrae, 1986).

There have been several studies that looked at some aspects of the Big Five and team performance. Very few, however, have looked at the effects of *Agreeableness* on team performance and none have looked at the relationship between these two variables as specifically and as in detail as was done in the current study. The following is what is currently known about personality and team performance.

Teams, Personality, and Computer Simulated Tasks

A team's average personality score and their achieved results is an area of Industrial/Organizational (I/O) psychology that is gaining momentum. This trend will continue as many corporations are increasingly emphasizing teamwork. Psychologists are working on computer simulations that should be able to give generalizable feedback that will translate into real world success. Several studies have used team performance on a computer simulated task and a personality index in order to look for relationship between the two. A portion of these studies like the one conducted by Colquitt, Hollenbeck, Ilgen, LePine, and Sheppard (2002), have met with success. The reason the Colquitt et al. (2002) study is being used as an example is that after an extensive computer literature search, no studies concentrating directly on *Agreeableness* and team performance on a computer simulated task could be found.

Colquitt et al. (2002) conducted their study with a computer simulation called the Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise

(TIDE²). They also incorporated Costa and McCrae's (1992) Revised Neuroticism, Extraversion, and Openness to Experience (NEO) personality inventory, with a specific emphasis on the aspect of Openness to Experience in order to look for a relationship between this personality dimension and team performance. They were successful in finding significant results between openness to experience and team performance ($\Delta \underline{r}^2 = .14$, $p < .05$) using the computer simulated task. For additional information on the TIDE², refer to Hollenbeck, Ilgen, LePine, Colquitt, and Hedlund (1998) or Gigone and Hastie (1997).

Another research team, Ellis, Hollenbeck, Ilgen, Porter, West, and Moon (2003) conducted a similar study to the one discussed above. Ellis et al. (2003) utilized the Distributed Dynamic Decision-making (DDD) computer simulation. The TIDE² discussed earlier is a derivative of the DDD. In their study, Ellis et al. (2003) focused on personality and team learning instead of team performance (The difference and relationship between team performance and team learning will be discussed later). For the Ellis et al. (2003) study, the experimenters defined team learning as "a relatively permanent change in the team's collective level of knowledge and skill produced by the shared experience of the team members" (p. 822). Among the variables examined in their study were the effects of Openness to Experience and *Agreeableness* on team learning. Ellis et al. (2003) noted that agreeable individuals tend to be compliant, self-effacing, modest, conforming, and non-confrontational, and though this may encourage team cohesion, but it may also detract from team learning. Team learning may also be hindered because if the group as a whole scores high on *Agreeableness*, they may be more likely to reach a premature consensus on a course of action. If a group reaches consensus prematurely, this may lead the team to overlook several significant steps of critical thinking that may have led to the team making a better decision. Their hypothesis on the relationship between *Agreeableness* and team learning is that: “project teams with higher levels of *Agreeableness* will evidence lower levels of team learning” (p. 823).

Results of the Ellis et al. (2003) study showed that *Agreeableness* correlated negatively with team learning on a computer simulation. This was consistent with Colquitt et al.’s (2002) findings on team performance mentioned earlier. In addition, Ellis et al. (2003) found that higher levels of Openness to Experience correlated with higher levels of team learning in a computer simulation. Both of Ellis et al.’s (2003) hypotheses regarding Openness to Experience and *Agreeableness* were supported.

A reason given by Ellis et al. (2003) for the negative relationship between *Agreeableness* and team learning is that premature consensus, due to a lack of conflict, has a detrimental effect on both problem solving and decision making in groups. This is part of a phenomenon known as group think. In the current study, it is believed that a group high in *Agreeableness* may come to a premature consensus regarding task that require critical thought to be executed properly. A definition used for group think is as follows: Group think is a phenomenon where alternative solutions to a problem are ignored due to an overassertive leader, an absence of diversity amongst opinions, or a group that has too much momentum in one direction (Janis & Mann, 1977). Group think often leads to a terrible final solution that can end up causing anywhere from a minor inconvenience to death, as was the case with the Challenger Shuttle accident. Janis and

Mann (1977) proposed a model on the phenomenon of groupthink that supported the finding of Ellis et al. (2003) by including a facet of "failure to re-examine preferred choice" (p.132) as something that may eventually lead to group think.

Moving from the concept of group think back to the concept of *Agreeableness* and teams, it is hypothesized in the current study that a curvilinear relationship exists between *Agreeableness* and team performance on a computer simulation. The reason for this hypothesis is that at the lower end of the *Agreeableness* spectrum teams will fail to agree on a solution where on the higher end, teams may agree too soon. In the next section, team learning and its relationship to team performance will be discussed since the concepts of team learning and team performance are so closely related. A relationship must be established between the two in order to get a better idea of why *Agreeableness* seems to improve team performance, but hinder team learning.

Relationship Between Team Learning and Team Performance

Team learning is defined as "a relatively permanent change in the team's collective level of knowledge and skill produced by the shared experience of the team members" (Ellis et. al. 2003, p. 822). Alternatively, team performance can have many definitions, but essentially refers to how well a team does on a given task. In theory it would seem that higher team learning would lead to better team performance, but in order to clarify confusion in the field between the relationship of team learning and team performance, Bunderson and Sutcliffe (2003) did a study of Fortune 100 companies that looked at this specific concept and found some unexpected results. Employing performance measures of actual profitability to targeted profitability (performance-to-plan) and actual profitability relative to units sold (profit-per-unit); Bunderson and Sutcliffe (2003) found that putting too much emphasis on learning may actually be deleterious to efficiency.

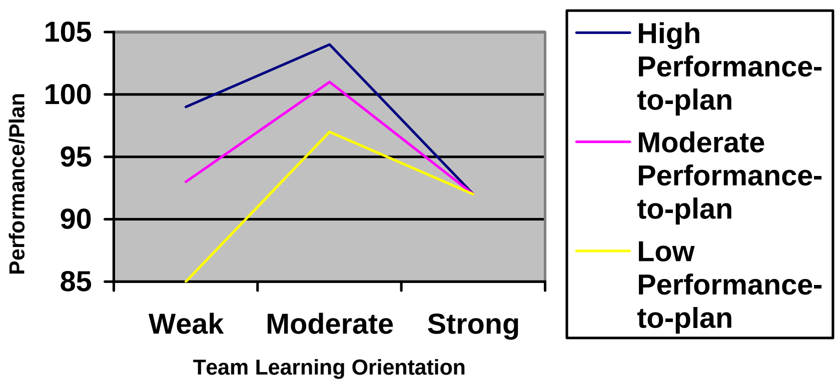

Figure 1. This is the predicted relationship between team learning orientation and business unit performance for different levels of team learning orientation.

This is especially true if the team over emphasizing team learning had previously been performing well. Though counter-intuitive, when the results are graphed using performance-to-plan and three levels of team learning orientation (weak, moderate, and strong), a curvilinear relationship emerged with performance peaking around the moderate area of team learning orientation, as shown in Figure 1.

These results suggest is that an overemphasis on team learning in a business setting may hinder performance. If one were to relate these findings to team *Agreeableness*, it might be hypothesized that an average level of *Agreeableness* for a team would foster both efficient team learning and team performance. If *Agreeableness* levels were to fall substantially below or above the average level, there may be a drop off in team performance. This leads back to the hypothesis of the current study that was stated earlier of *Agreeableness* having a curvilinear relationship with team performance. In the next section *Agreeableness* and a shared mental model, which is important to successful team performance, will be discussed. A shared mental model essentially means that each team member has a similar picture in his or her mind of the information available and the best way to go about solving the given task.

Agreeableness and a Shared Mental Model

In a 1999 study by Neuman and Wright, it was determined that *Agreeableness* should help a group come to a consensus on a shared mental model (SMM), which they defined as a "group conceptualization of the environment and how to interpret it that transcends the cognitive approaches of the individual" (p. 379). Another interesting finding of the Neuman and Wright (1999) study was that *Agreeableness* in teams and supervisor task ratings (also referred to as task performance in the study) were positively correlated, with $r = .36$, ($p<.01$). This is in contrast to the findings of the Ellis et al. (2003) findings that *Agreeableness* is detrimental to team performance. Another finding in contrast to the study done by Ellis et al. (2003) was that *Agreeable* individuals were more likely to be effective in group activities requiring coordination between the group members. Some of the reasons given by Neuman and Wright (1999) for their findings were that group members high in *Agreeableness* are better at avoiding disruptions at work that might be brought about by interpersonal conflict. They also mentioned that

Agreeableness should help a group come to a consensus on a SMM. It may be harder to draw definite conclusions from this study due to the subjective nature of supervisor task rating; however, this rating system may be more applicable to team work in organizations. Neuman and Wright (1999) went on to discuss the different facets of *Agreeableness* in their study (i.e. trust, straightforwardness, altruism, compliance, modesty, and tender-mindedness). They stated that one would expect tender-mindedness, altruism, and trust to enhance interpersonal skills, thus allowing organization members to relate effectively to others. Neuman and Wright (1999) went on to mention that compliance and straightforwardness should indicate sincerity in an individual and also signify willingness to work towards productive information-seeking and negotiation tactics. They did not indicate how the facet of modesty would relate, but one might posit that modesty would facilitate a team performing well. Unfortunately, Neuman and Wright (1999) only tested *Agreeableness* as a construct in their study and not any of the individual facets that make up *Agreeableness*. In the current study, the individual facets of *Agreeableness* and their effects on team performance on a computer simulated task are hypothesized and tested.

Other studies that had similar results to Neuman and Wright (1999) are Bennet and Carbonari (1976) and Kilmann and Thomas (1975). These studies found that teams high in *Agreeableness* will have an easier time agreeing upon a shared mental model. In contrast to this opinion, the experimenters of the current study would argue that if a team is too collectively agreeable, then there may be an absence of the conflict that might stimulate the formation of the most optimal SMM. In the opposite direction, a team very low in collective *Agreeableness* may also struggle to form the optimal SMM because if one or more members of the team are unwilling to share their knowledge or information on a particular subject area important to the team, the team will be unable to make the best and most accurate decisions regarding a course of action (Klimoski & Mohammed, 1994).

Teams, Personality, and "Back Up Behavior"

McIntyre and Salas (1995), through their studies on teamwork, determined that there were four essential aspects to teamwork. These four aspects are Backing Up Behaviors, Closed-loop Communication, Performance Monitoring, and Feedback. We will concentrate on the concept of Back Up Behavior in this section. Porter, Hollenbeck, Ilgen, Ellis, West, and Moon (2003) revisited the concept of backing up behaviors and included the Five Factor Model (FFM) personality types in a more recent study. Porter et al. (2003) returned to this concept of backing up behavior because they felt backing up behavior may be the most critical aspect of teamwork given by McIntyre and Salas (1995). Backing Up Behaviors, as defined by Porter et al. (2003) means that team members will help each other to perform the task on which they are currently working. Some examples given are correcting the mistakes of a fellow team member, or if a team member is unable to perform a certain task assigned to him or her, another team member will step in and fulfill the duty. The specific definition used for the Porter et al. (2003) study was "the discretionary provision of resources and task-related effort to another member of one's team that is intended to help that team member obtain the goals as defined by his or her role when it is apparent that the team member is failing to reach those goals" (p. 391).

The task that the participants were asked to perform in the Porter et al. (2003) study was set up using the DDD. This particular variation of the DDD involved coordination of several military vehicles using Airborne Warning and Control Systems (AWACS). In the Porter et al. (2003) study, team members were stationed in a common room in close proximity and used networked computers, which is comparable to the current study. Teams in the study consisted of four individuals and in total there were 71 teams. In order to facilitate backing up behaviors, one team member out of the four (designated DM2, or Decision Maker 2), was given a disproportionately heavy workload compared to the other three team members. Porter et al. (2003) did not specify whether or not this person was assigned randomly. As was done in the current study, Porter et al. (2003) administered the personality test to the participants before the DDD task was performed. They looked at all five of the FFM personality types in the study, but specifically and pertinent to the present study was their hypothesis that states, "in teams, provider *Agreeableness* and legitimacy of need will interact in determining the amount of back up behaviors."(p. 395)

The result of the Porter et al. (2003) study regarding *Agreeableness* was not what was expected. They found no significant effects for *Agreeableness* and legitimacy of need, which seems counter-intuitive. They did test for all six of *Agreeableness* facets, but only one, altruism, showed any effect regarding legitimacy of need. The relationship was in the expected direction, with higher altruism being associated with higher back up behaviors. A possible explanation given by the authors is that perhaps the *Agreeableness* was indiscriminate in nature and applied in more of a blanket approach, rather than being directed at the team member who truly needed the most help by being backed up, but upon post hoc examination, no evidence for this theory was provided. We do not specifically look at Back Up Behavior in this study, but future studies we perform on this subject will investigate this.

Why use a Computer Simulation for a Team Task?

The following section is dedicated specifically to computer simulations and their use for assessing team performance. Specifically, the current study used a computer simulation called the Distributed Dynamic Decision-making (DDD) task. This allowed the experimenters to place the participants in an environment that they were probably not familiar with and test how well they did. Using a task or environment with which most people are not familiar is usually chosen in order to test specific abilities in a laboratory. This is done so that there is little chance of outside practice effects becoming a nuisance variable. It should be noted that it is possible to have computer simulations that are designed to mimic a work place such as an office or a warehouse in order to better train employees. It should be noted that because a computer simulation is being used for experimental purposes and not training in the current study, the location of the Arctic was chosen instead of a familiar locale. This served to put the participants in an equally unfamiliar environment.

Specifically, the current study falls into a category referred to as Computer Simulations for Team Research (CSTR). According to Rogleberg (1999), only about 10% of the studies done between 1996 and 1998 on team work utilized computer simulations. This will increase in the coming years as technology allows more complex situations to be accurately modeled. In addition, the interfaces will become easier for participants to use and will only require a minimum amount of experience and training to be used.

Marks (2000) pointed out that there are two types of computer simulations being used in teamwork research today. Simulations can be defined as situations created in order to "place individuals in complex, dynamic, and malleable situations not easily created" (p. 655). Simulations are critical for modeling real world team problems because they allow real life situations to be recreated without creating danger to participants or equipment (Schiflett, Elliot, Salas, and Coovert, 2004). The first type of computer simulation is a task modeled on a real world situation. Examples of this could be a flight simulation for pilots, a tank simulation for a tank team, or a stock market simulation for stock brokers. With these simulations, one can introduce predicaments such as an engine flame out on a jet, loss of night vision in a tank, or a stock market crash, without violating the principles of ethics or endangering anyone. These specific simulations also allow a realistic crisis to be created without spending exorbitant amounts of money on the actual hardware that would be required to obtain the same level of training results without a simulation. There is another type of computer simulation that differs from the first in that it does create a realistic environment; however it is supposed to be unfamiliar to participants in order to remove practice effects or other type of effects due to previous experience.

The second type of computer simulation, and the type that will be used for the current study, is referred to as a Hypothesized Nomological Net (Marks, 2000). This type of simulation is used to test the relationships that exist within a team under a variety of situations. In the current study, Arctic Survival was chosen because it will most likely be unfamiliar to the participants, but still be able to stimulate critical thinking and allow assessment of how they work together as a team.

For a computer simulation to be a valid way of assessing team performance, Raser (1969) laid out the following four criteria. The simulation must have psychological reality, process validity, structural validity, and predictive validity. For psychological reality to be present the participants must believe that they are part of a real team: additionally, for the task(s) to be completed successfully, the team must depend on each other and work together. If the participants do not believe that they are a part of a real team and that success requires collaboration, then the results will not be valid. We attempted to induce this into the study by training the participants together and offering a cash prize to the team who obtained the highest score on the task. The other types of validity process, structural, and predictive validity, must be induced differently.

Process validity is achieved when the process one is attempting to test is present in the simulation. It is not possible to simulate all the real world details in a simulation but whichever facets are being tested must be present in the simulation. In our study we are looking at critical thinking and collaboration. These are both necessary to successfully complete the Arctic survival computer simulation. This leads to structural validity.

Structural validity is made possible by accurately representing the configuration of the real world or Nomological-net model as it would be in reality. An example of the real world model having structural validity would be as follows: if one is trying to simulate an actual cockpit crew in a commercial jetliner, one would have to simulate the responsibilities of the pilot, co-pilot, and navigator. If all three participants have control of the planes control surfaces (i.e. rudders, elevators, ailerons, etc.) and navigation equipment, then structural validity has been lost. If, in the case of a Nomological-net simulation, one is looking at leadership decisions it is necessary to create a hierarchical team-member structure in addition to making the proper information and resources accessible to the leader (Marks, 2000). In the simulation used in the current study, each member only had access to a certain amount of limited resources, thus guaranteeing interdependence and cooperation in order to succeed. This leaves us needing to satisfy the criteria for predictive validity in order to realistically asses team performance.

Predictive validity alludes to the simulation predicting the relationship that occurs in the reference system, which means that if a relationship is known to occur in reality, then it should also occur in the simulation (Marks, 2000). An example given by Marks (2000) is that if a business simulation indicates that strategic planning enhances team confidence levels, and one knows that these constructs are related through existing research, then this would present evidence for the predictive validity of that particular simulation. It is very difficult to generalize simulations of this specificity and it is important to validate each individual simulation to a particular situation. In the current study, participants must cooperate and communicate in order to be successful and obtain a high score. Therefore, the criteria for predictive validity are satisfied.

In addition to the findings in the Marks (2000) study, a study done by Thompson and Coovert (2002) mentions that using computers for a team task can influence many team decision processes, such as conforming to team norms (which working on the computer lessens) and equalizing the amount of participation by each group member.

The Current Study

So far we have discussed the history of personality and how we have arrived at the current number of five factors. We have also discussed what the five factors are and of what each one is comprised in order to distinguish it from the other factors. The specific focus of the personality discussion was based on *Agreeableness* because it and its six factors will be a main focus of the current study.

Team performance and team learning were compared and contrasted and then two studies which had looked into both team performance and team learning using a computer simulation similar to the one being utilized in the current study were mentioned. In addition, the history and different types of computer simulations were discussed and analyzed. The two different types of computer simulations and the four types of validity that a computer simulation must possess in order to be considered valid were presented.

Also discussed were previous studies that involved teams, personality, and computers and their findings. Gaps in the literature were noted and conflicting results of similar studies were brought to light. The current study attempts to fill in some of the gaps in the literature and resolve the conflicting results of previous studies or perhaps push the theory towards a certain direction. The following paragraphs discuss the current study's theory and methods in detail.

Data for the current study were archival. The original data came from a study that looked at how either collaborative critical thinking (CCT) or survival training and either a presence or absence of information probes during the task affected team performance. The experimenters also had the participant take a personality survey based on Goldberg's (1999) International Personality Inventory Pool (IPIP) with 60 of the 100 items coming from the item pool for *Agreeableness.* It was the administration of the *Agreeableness* survey in the previous study that made the current study possible.

In the current study we tried to determine the relationship between the personality trait of *Agreeableness* and its effect on team performance. Previous studies found mixed results, which is why the current study, with a specific focus on *Agreeableness* and team performance, has been proposed. Participants were placed into teams of three (two participants and one administrator) and given a computer simulated task to perform as a team and their results were objectively scored to determine how they performed. Specifically, the computer simulation was the Distributed Dynamic Decision-making (DDD) task, which has been shown to be valid in several previous studies (Colquitt et al., 2002 & Ellis et al., 2003). This study utilized the Arctic Survival version of the DDD. The *Agreeableness* of the participants was measured using the IPIP. In addition, the six facets of *Agreeableness* were individually assessed. Our hypotheses for the study are listed in the following section.

Hypotheses

Following the discussion of how *Agreeableness* and its specific facets will affect team performance, these are the hypotheses of the current study.

Hypothesis 1. Teams whose average score is higher or lower than the mean on Agreeableness will perform worse in terms of cumulative team points scored than teams who score around the mean on the Agreeableness scale, therefore resulting in a curvilinear relationship.

Due to the nature of personality an exact mean number for the *Agreeableness* score will not be known until after the completion of the study. Once the mean of the *Agreeableness* scores for all the participants is calculated, then a mean number for *Agreeableness* can be assigned. It is surmised that teams with a higher or lower level of *Agreeableness* than the mean will have a lower performance score on the task than teams who are closer to the mean, because members of teams who are more agreeable, while being less combative may also be less critical of each other. Because they are less critical of each other, this may lead to more errors or a failure to perform to their optimal abilities. For the teams scoring lower than the mean on the *Agreeableness* scale, an inability to coordinate or agree will result in poorer performance on the task.

In addition, it is hypothesized that the relationship with the six facets of *Agreeableness* will be as follows. These hypotheses are based on Ellis et al.'s (2003) study in which a negative relationship between team learning and *Agreeableness* was found. In addition, these hypotheses are based on Neuman and Wright's (1999) study where a positive relationship between *Agreeableness* and task performance was found and Bunderson and Sutcliffe's (2003) study where more team learning did not lead to better team performance. The hypotheses on the facets of *Agreeableness* are as follows:

Hypothesis 2. Trust will correlate positively with team performance because if the team members trust each other they will be more likely to cooperate and collaborate, therefore scoring higher.

Hypothesis 3. Morality will correlate positively with team performance since treating one's teammate fairly and not withholding information or resources will be advantageous to the team.

Hypothesis 4. Altruism will correlate positively with team performance because if one tries to do all the tasks alone then the score will suffer. Teams must be willing to share in the tasks equally or some will be left untended.

Hypothesis 5. Cooperation will correlate positively with team performance because working together is part of being a good team.

Hypothesis 6. Modesty will correlate negatively with team performance because if one or both team members are satisfied with scoring only a modest amount, then time and resources may be wasted tending to unimportant tasks.

Hypothesis 7. Sympathy will correlate negatively with team performance because if one or both team members are too concerned with giving orders that may lead to important tasks being uncompleted, and time and resources may be wasted tending to unimportant tasks.

These hypotheses are partially based upon Neuman and Wright's (1999) untested speculations on how the six facets of *Agreeableness* would individually affect a team. No studies could be found where the individual facets of *Agreeableness* were examined to determine if any or all of them enhanced team performance.

In addition to the hypotheses on *Agreeableness*, the final hypothesis will be based on general mental ability (GMA) and team performance. This was included to try and replicate the results of previous studies that found a link between GMA and team performance such as Neuman and Wright (1999) among others. Our hypothesis is as follows:

Hypothesis 8. *Teams with higher GMA will score higher than teams with lower GMA.*

Method

Participants

In this study we had 144 undergraduate psychology students from a large university in the southeast divided into 2 person teams (72 total teams), who performed a task on the Distributed Dynamic Decision-making (DDD) computer simulation. Of these 72 teams, 62 teams provided us with usable data. Age and ethnicity should not be a factor as our sample is assumed to be representative of the overall undergraduate population. Gender, however, may be a problem as our sample was comprised of 81% females. Although the exact effects of gender and teams are currently unknown, the gender makeup of the participant pool did not represent the overall population. However, it did represent the population of the psychology department at the university where the experiment was conducted.

Materials

Two computers configured to run the DDD were used to administer the task to the participants. Before participants began the task they viewed a series of Audio Video Interleaves (AVIs) running on a computer that demonstrated how to utilize the tools of the DDD. Several handouts that will be mentioned more specifically in the procedure were also utilized.

Procedure

Each experimental session had a running time of three and a half hours with the actual DDD Arctic Survival task the participants were scored on running for 75 minutes. When the participants came into the lab, they were instructed to sit at the round table in the center of the lab. After brief introductions, participants filled out informed consent forms and demographic sheets. Once this was finished, participants were instructed to draw a plastic tab out of a bowl held above the participants' head. The tabs were marked either A or B and were used to determine which station the participant would sit at for the DDD task. After the participants drew the tabs, we had them sit in front of a computer where the administrator gave them a power point presentation on either Survival Training or Collaborative Critical Thinking (CCT) training.

After the training presentations were finished we had the participants return to the round table in the middle of the lab and we instructed them to build a tower out of the tinker toys we provided. This exercise was done in order to facilitate the participants becoming more familiar with each other. The tinker toy task lasted for ten minutes and in order to make this task more interdependent for the participants, we instructed person A to only touch the joiners and person B to only touch the sticks. This way a tower could not be built by one person alone. Once the tower building phase of the experiment was over, the tinker toys were put away and the participants then filled out the personality questionnaire which will be discussed later (Appendix A).

After each of the participants had finished the personality questionnaire, we instructed them to sit at a computer located in the middle of the room where they viewed the AVIs we had prepared. The AVI session was divided into thirds and encompassed all of the necessary training needed for the DDD. After each of the three sessions was over, we would give each team member five to seven minutes to utilize what they had learned on a practice DDD scenario. This scenario was similar to the DDD scenario they would do for the experiment. In total, each team member received fifteen to twenty one minutes of actual time on the DDD in addition to the forty five minutes of AVI training. Once the last session of hands on training was complete and all of the participants' questions about the DDD had been answered, the participants went back to the center table and the computers were set up to run the experimental DDD scenario. Teams were given a background information page (Appendix B) to brief them on what had theoretically taken place in the Arctic Survival scenario before they arrived. The participants were instructed that the goal of the task was to locate an unmanned aerial vehicle (UAV) and a lost team who was stranded. They were given a sheet with basic strategies and the point values for various tasks they could perform in the scenario. This was titled "tactical information" and is located in Appendix C. The DDD Arctic Survival scenario is explained in more detail in the next section.

In the Arctic Survival scenario there are four separately color coded team members. They are the red snow cat, the purple snow cat, an observer coded as green, and the blue fuel cat. The administrator played the blue fuel cat. The fuel cat only reacted to orders given by the participants. The green team member had the ability to see the entire map and everything that was going on in real time on the map. The reason for including the green team member in a scenario that only required three people was that the DDD Arctic Survival scenario was originally programmed to use four people, three snow cats and one fuel cat. Since we were only using two participants and one administrator, the green team member was reduced to an observer. If there had been an option to omit the green team member from the scenario that is the course of action we would have taken. To reiterate, the green team member did not have a way to participate in the game and was solely an observer, the DDD has the ability to run with three snow cats and a fuel cat, but for this scenario the teams consisted of only two snow cats and a fuel cat. Green was fixed as an observer and no green snow cat icon was visible during the scenario. The red and purple participants did not know about the green team member.

The blue team member or fuel cat, which was controlled by the administrator, only had the ability to refuel the red and purple team members. The administrator only used the blue fuel cat to refuel the red and purple team member when either the red or purple team member indicated that they needed refueling by emailing the blue fuel cat and enabling their refueling icon. The refueling icon could be seen by the blue team member if the blue fuel cat was within sensor range of the current snow cat in need of refueling. In addition, the participants were instructed to direct any questions they had during the running of the DDD to the blue fuel cat.

The red and purple team members, who were controlled by participants, were the only snow cats able to complete the different tasks of the scenario. Each red and purple team consisted of a snow cat, a medic, a technician, a scout, and a mechanic. All four of the personnel could be put onto a snow cat and transported to various locations. In addition, each of the personnel (medic, technician, scout, and mechanic) starts out with a certain amount of usable units. For example, the medic starts out with 15 medic units. If a task requires 3 medical units to complete and the red medic is assigned to a task that indicates it needs 3 medical units, then after completing the task, the red medic will have 12 medical units. This is similar for each of the other four color coded team members who each start with a certain amount of points. Each team, red and purple, had the same personnel and each of these personnel had the same amount of points as their counterparts on the other team (i.e. red technician had 15 units, so purple technician had 15 units). Communication between the red and purple participants and the blue administrator was only done electronically through a messenger system built into the DDD.

Scoring for the DDD was recorded automatically by the computer and is explained in more detail below. In addition to receiving points for finding the crashed UAV or the lost team as they had been instructed to do, the participants could score points by completing such tasks as fixing a rusty drill, administering non-emergency medical assistance, or an assortment of other activities.

Scoring:

Scoring was divided into three different areas. Objective scoring was accomplished by simply looking at the scores each team receives in accordance with the DDD. Each team had an opportunity to score points on their tasks. The point system is as follows:

Point Allocation:

300 points: Find the unmanned aerial vehicle (UAV) or the lost team.

100 points: Render Emergency assistance (-100 from both team members if emergency assistance is not rendered in the allotted time period)

50 points: Assist with repair or medical requests

10 - 80 points: Process seismic monitors

Some of these tasks were available at the beginning of the scenario and others popped up at predetermined intervals and were relayed to the red and purple snow cats via the blue fuel cat played by the administrator. The participants were told by the satellite messages relayed from the blue fuel cat where certain tasks were located. If the participants successfully completed a task, they were awarded the above amount of points, depending on the task. There was no time limit to completing the tasks, with one key exception, which were emergencies.

Emergency assistance had to be rendered within 15 minutes of when the participants received the e-mail alerting them to the emergency. If the participants successfully administered emergency assistance in the given time, then they scored 100 points. If however, they were unsuccessful, they lost 100 points each. An anomaly in the programming for the DDD was that an emergency could be neutralized by processing it with only part of the needed resources (i.e. the emergency requires 3 medical units, but the participant who attends to the emergency only has 2 medical units). In this case the team was neither penalized nor rewarded for attending to that particular emergency and received 0 points. To clarify, if a team had a combined 300 points and one of the teammates attended to an emergency that required 3 medical units with only 2 medical units, they would still have 300 combined points afterwards instead of the 200 they would have had if they missed the emergency or the 400 they would have had if they had successfully attended to the emergency. Fortunately, this anomaly is believed to have happened only one time.

In order to determine a team's average level of *Agreeableness*, each individual was administered the *Agreeableness* portion of the International Personality Inventory Pool (IPIP) (Goldberg, 1999) before his or her training on the DDD (Appendix A). The IPIP consists of 60 *Agreeableness* questions with 10 questions from each of the six facets of *Agreeableness*. The other 40 questions were comprised of the other four areas of the IPIP. The questions from the different factors were randomized in order to prevent priming for *Agreeableness*, as the IPIP was administered before the task. The two individual's scores on the IPIP were then combined and averaged to form an average team *Agreeableness* score. In addition to analyzing the average *Agreeableness* score, the scores of the individuals of each team were analyzed in order to determine if the variance within a team had any effect on team performance.

To assess GMA, an area for SAT/ACT score was included on the demographics sheet. If a participant was unsure of his or her SAT/ACT score, they were instructed to estimate it. In the case of estimation by a participant, a note was made on the demographics sheet. If the participant had never taken either of the tests or was unable to estimate their score, the space was left blank. After the study was concluded a formula was used to transform all SAT scores into an equivalent ACT score and all data used in the results section is based of the transformed scores.

Results

Data Analysis

All tables referred to in this section can be located at the beginning of the document or located through the table of contents. In our population of 62 teams, there were 24 males (19%) and 100 females (81%). Table 1 lists the descriptive statistics for the study. Team scores ranged from -400 points (teams who missed all four emergencies and completed no tasks) to 1420 points.

Table 1
Descriptive statistics

Variable	Mean	Standard Error
Age		
Male	20.46	.643
Female	21.69	.382
SAT/ACT conversion to ACT		
Male	24.43	.628
Female	24.37	.460

An ANOVA was performed on the experimenters and the scores of the teams they administered the task to and was not significant ($F(4,61)= 1.030$, $p=.400$) meaning that there was no experimenter effect (Table 2)

Table 2
ANOVA performed on the average team score and grouped by experimenter

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	659371.703	4	164842.926	1.030	.400
Within Groups	9123049.265	57	160053.496		
Total	9782420.968	61			

Table 3 shows average team score, average level of *Agreeableness*, and the average level of the six facets of *Agreeableness*.

Table 3
Mean and standard deviation of team score, level of Agreeableness, and level of the six facets of Agreeableness

Variable	Mean	Standard Error
Team score	311.129	394.051
Agreeableness	2.91	.18
Trust	3.11	.15
Morality	2.29	.2
Altruism	3.45	.22
Cooperation	2.35	.25
Modesty	3.45	.15
Sympathy	2.90	.1

The original *Hypothesis 1* stated that a curvilinear relationship between team average team *Agreeableness* and team score would be explored. The mean of average team *Agreeableness* was 2.91 (*SE*=.18) and the range was only 1.483 with a small standard error. This indicated that there was a possible range restriction for the responses to the questions due to the Likert scale ranging from only one to five. Due to the small amount of variability and small range in the team *Agreeableness* scores, a curvilinear analysis was deemed unfeasible. In addition, a linear regression analysis was also deemed unfeasible because we had individuals nested within teams. Therefore, the program MLWin and the method of Hierarchical Linear Modeling (HLM) were used for the analysis of *Hypothesis 1.* As noted before, individuals were nested within teams and in addition to *Agreeableness*, some of the other level-1 variables were used to serve as predictors in the analysis were Intelligence Quotient (IQ), which was ascertained via SAT and ACT scores. Age, individual score on the task, gender and year in college were the other variables included. The level-2 variable was the team to which the individuals belonged. Table 4 shows the different models analyzed using HLM to predict team score.

Table 4
Model Comparisons of data using HLM to predict team score

Model	Variables	Overall Deviance (χ^2)	Change in Deviance	Prob.
A.	Null	1836.05		
B.	Agreeableness	1834.01	2.04	Ns
C.	IQ	1266.97	569.08	.001
D.	Individual score	1680.7	155.35	.001
E.	IQ & Agreeableness Compare to agr. Compare to IQ	1265.3	570.75 568.71 1.67	.001 .001 ns
F.	Intercept and Condition	1816.88	19.17	.001

Table 5 includes HLM models predicting individual score and was included for comparison.

Table 5
Model Comparisons of data using HLM to predict individual score

Model	Variables	Overall Deviance (χ^2)	Change in Deviance	Prob.
A.	Null	1751.83		
B.	Agreeableness	1733.77	18.06	.001
C.	IQ	1201.59	550.24	.001
D.	IQ & Agreeableness	1200.38	551.45	.001
E.	Intercept and Condition	1720.96	30.87	.001

Table 6 includes significant β weights from the HLM equations.

Table 6
Beta weights associated with different variables in the HLM model for team score

Variable	β	Ω	β/Ω (Sig. if β/Ω>2)
Agreeableness	206.6	144.270	1.43
IQ	22.93	10.97	2.09

For *Hypotheses 2* through *7*, linear regression was used to analyze the data and the results along with those for *Hypothesis 1* are summarized below.

Test of Hypotheses

Table 7 summarizes the results of the linear regression test of the hypotheses on the different facets of agreeableness.

Table 7
Summary of linear regression for Agreeableness Facets predicting team score (N=62 teams)

Variable	Adjusted r^2	B	SE B	β	Sig.
Average Trust	-.008	152.007	213.473	.092	.479
Average Morality	-.003	-154.731	173.765	-.114	.377
Average Altruism	.011	-297.186	229.384	-.165	.200
Average Cooperation	-.008	-114.116	158.268	-.093	.474
Average Modesty	.000	232.458	230.936	.129	.318
Average Sympathy	-.017	-4.571	207.829	-.003	.983

Discussed below are the general findings on the facets.

Hypothesis 1. This hypothesis stated that teams whose average score is higher or lower than the mean on *Agreeableness* will perform worse in terms of cumulative team points scored than teams who score around the mean on the *Agreeableness* scale. Our results showed that *Agreeableness* does not affect how a team performs in terms of points scored on a computer simulated task (χ^2 =2.04, p>.05)., but *Agreeableness* does affect how an individual performs on the task (χ^2 = 18.06, p=.001).

Hypothesis 2. Trust will correlate positively with team performance because if the team members trust each other they will be more likely to cooperate and collaborate and therefore score higher. The relationship between average level of trust and team score was not significant (*F 1, 61=.021, p=.884*).

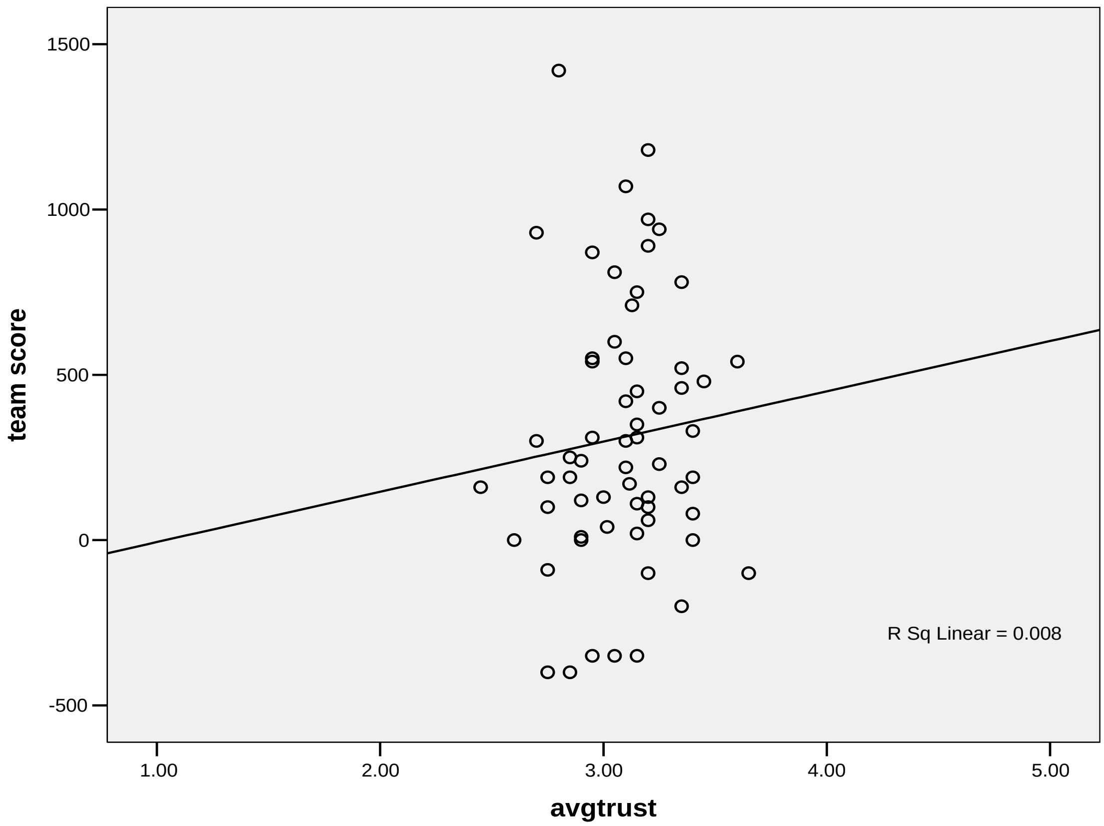

Figure 2. Scatter plot of average team level trust and team score with the best fit line.

Hypothesis 3. Morality will correlate positively with team performance since communication is limited to e-mail only and this being a timed task, succinctness and being direct will be advantageous to the team. The relationship between average level of morality and team score was not significant (*F 1, 61=1.747, p=.191*).

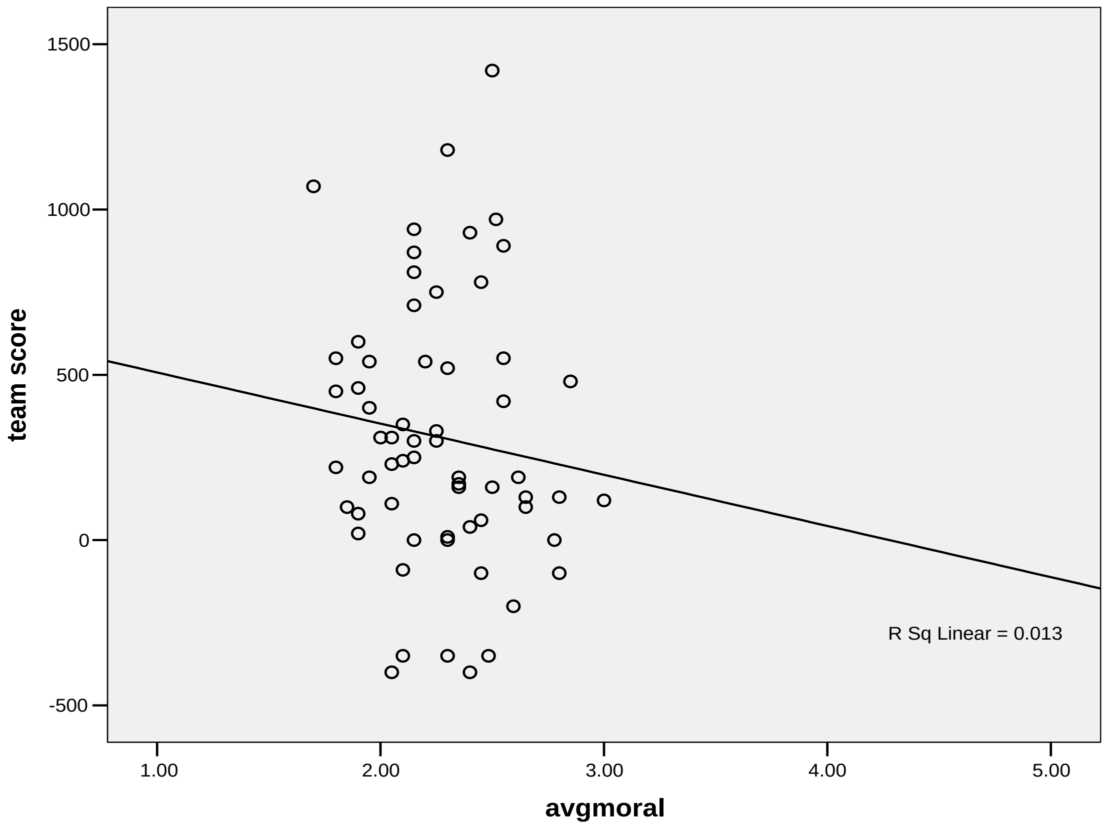

Figure 3. Scatter plot of average team level morality and team score with the best fit line.

Hypothesis 4. Altruism will correlate positively with team performance because if one tries to do all the tasks alone then the score will suffer. Teams must be willing to share in the tasks equally or some will be left untended. The relationship between average level of altruism and team score was not significant (*F 1, 61=2.947, p=.091*).

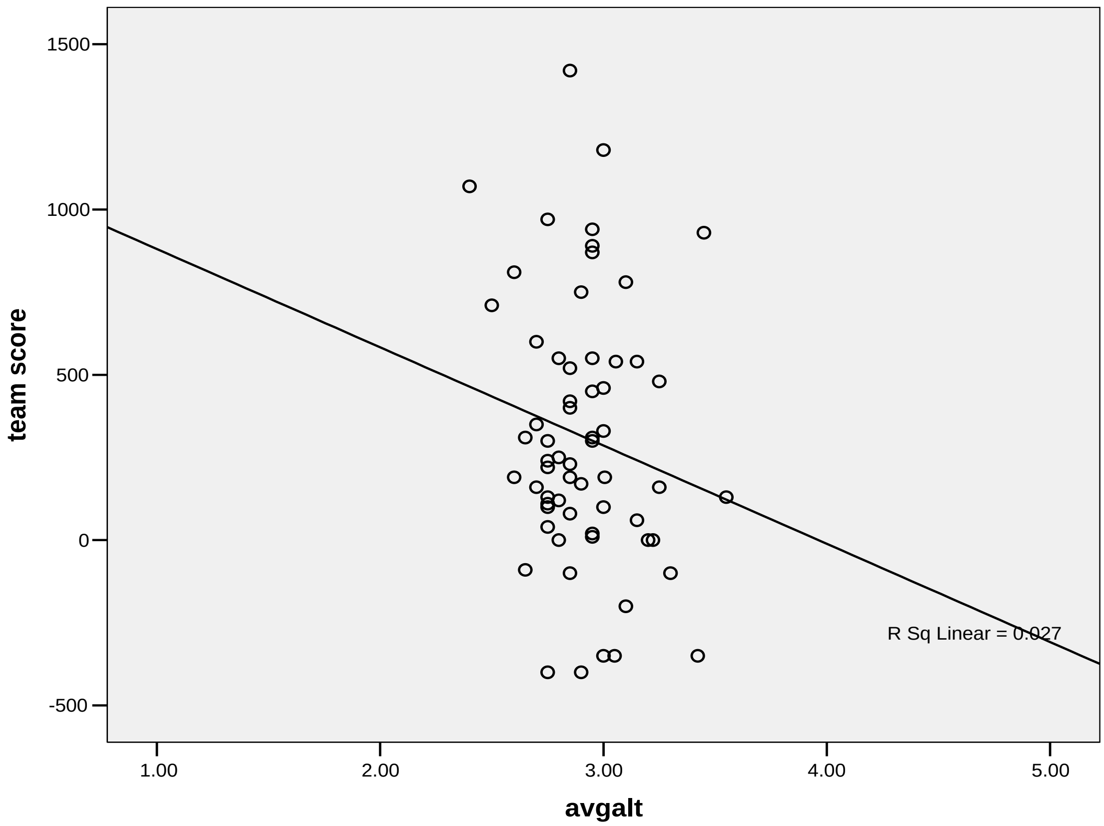

Figure 4. Scatter plot of average team level altruism and team score with the best fit line.

Hypothesis 5. Cooperation will correlate positively with team performance because working together is part of being a good team. The relationship between average level of cooperation and team score was not significant (*F 1, 61=1.71, p=.196*).

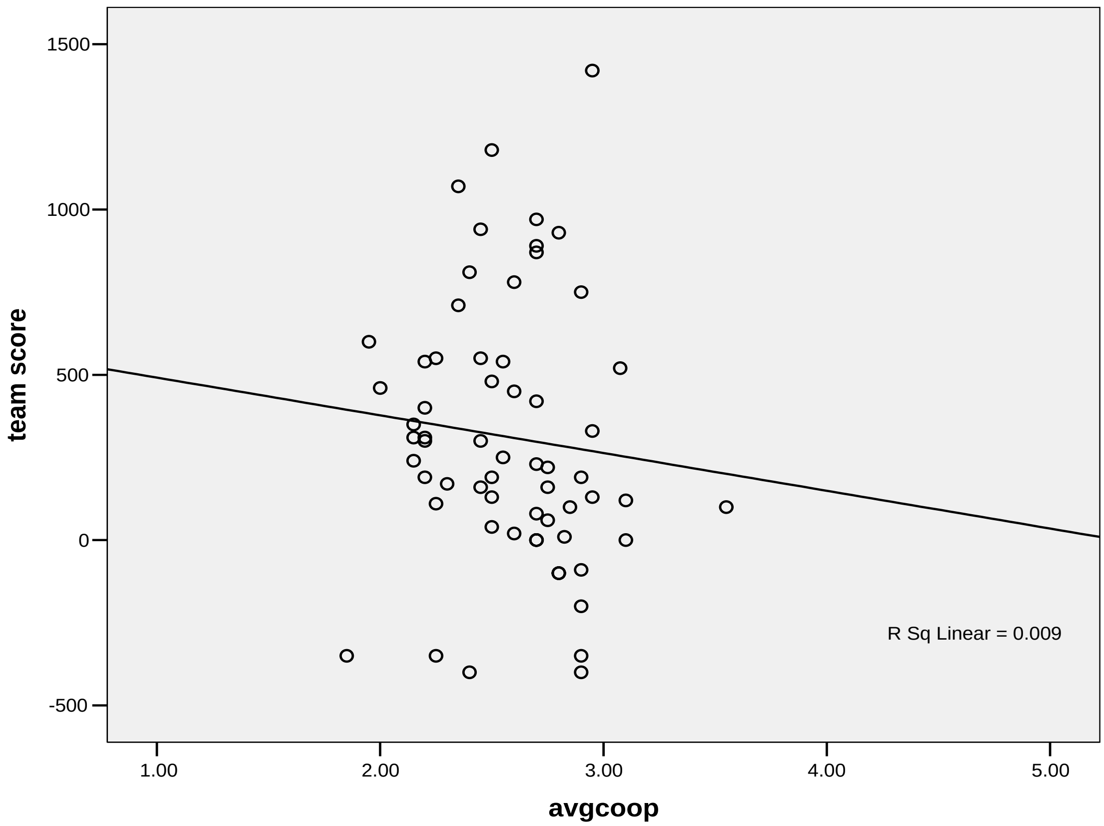

Figure 5. *Scatter plot of average team level cooperation and team score with the best fit line.*

Hypothesis 6. Modesty will correlate negatively with team performance because if one or both team members are satisfied with scoring only a modest amount then time and resources may be wasted tending to unimportant tasks. The relationship between average level of modesty and team score was not significant (*F 1, 61=.000, p=.985*).

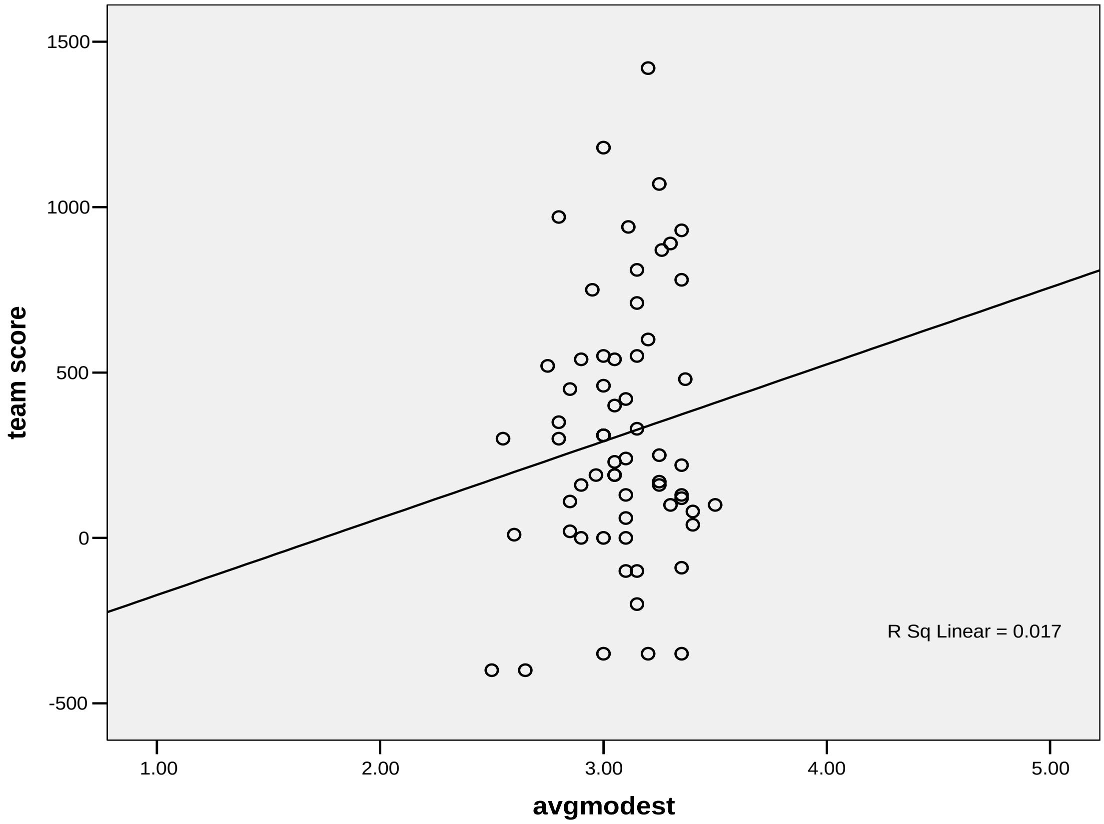

Figure 6. *Scatter plot of average team level modesty and team score with the best fit line.*

Hypothesis 7. Sympathy will correlate negatively with team performance because if one or both team members are too concerned with giving orders that may lead to important tasks being completed, then time and resources may be wasted tending to unimportant tasks. The relationship between average level of sympathy and team score was not significant (*F 1, 61=.488, p=.488*).

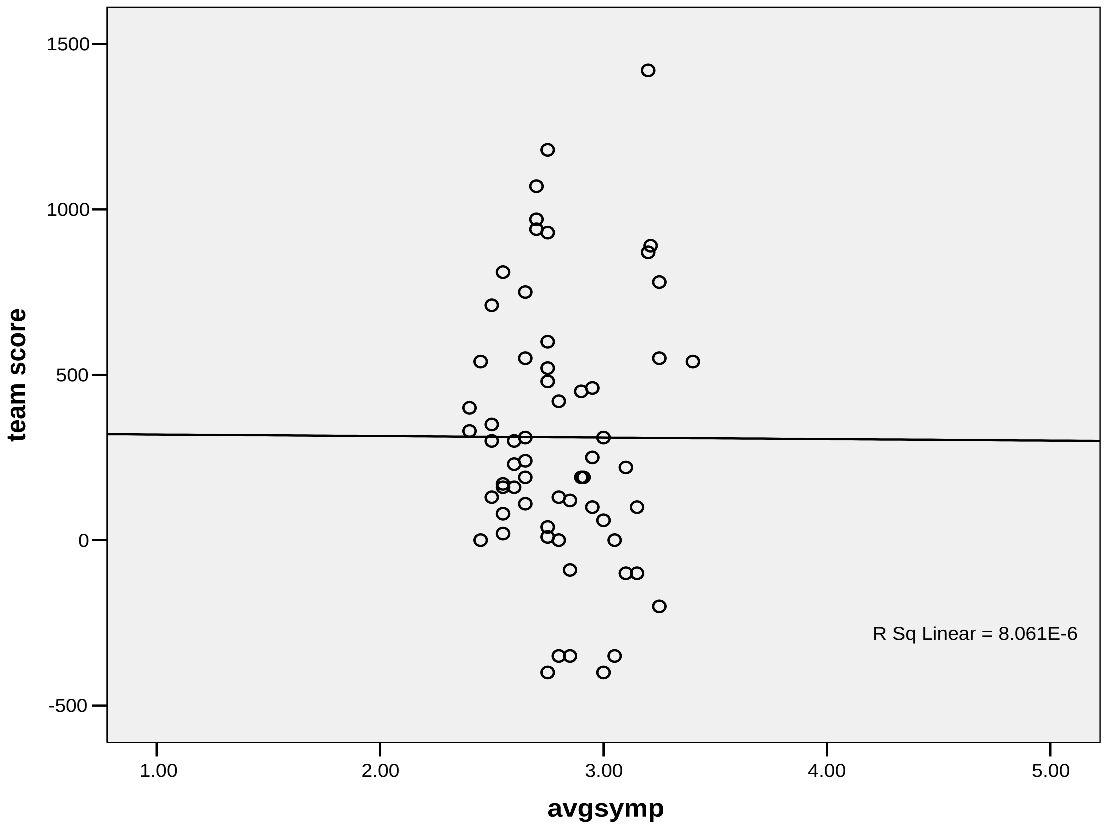

Figure 7. Scatter plot of average team level sympathy and team score with the best fit line.

Hypothesis 8. Teams with higher GMA will score higher than teams with lower GMA. This, like team and individual *Agreeableness,* was computing using HLM and this hypothesis was supported (χ^2 =569.08, p=.001).

Power Analysis

Determining power for a multilevel model is a complex process that is still lacking a definitive method of determination. However, there is a general consensus among the researchers using multilevel models that even with a sample size of around 30, an estimate of residual error at the lowest level (level-1), is still very accurate (Hox & Maas, 2002, Barcikowski, 1981). Recall for this experiment, that our sample size was n=124.

Results for training conditions from the archival study

The data from the type of training we were trying to prime, CCT or Survival, was not directly related to this study, but was a main focus of the archival study that we drew the data from to do our analysis. There were no hypotheses directly related to the training data in this study, but we decided to present the results in Table 8 in order for other researchers to draw their own conclusions about how the training data may relate to the hypotheses in this study. In Table 8, it is evident that there is a significant difference in scores between the five different conditions, and the post hoc test indicates that this significant difference is between condition 1 (no CCT or Survival training, no probes) and condition 5 (survival training, probes.)

Table 8
ANOVA and Post Hoc test on CCT and Survival training conditions.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1531226.126	4	382806.532	2.644	.043
Within Groups	8251194.841	57	144757.804		
Total	9782420.968	61			

Dependent Variable: Score
Tukey HSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-42.714	146.181	.998	-454.51	369.08
	3	-214.714	146.181	.587	-626.51	197.08
	4	-94.603	151.583	.971	-521.61	332.41
	5	-421.548(*)	137.682	.027	-809.40	-33.70
2	1	42.714	146.181	.998	-369.08	454.51
	3	-172.000	170.152	.849	-651.32	307.32
	4	-51.889	174.814	.998	-544.34	440.56
	5	-378.833	162.908	.152	-837.75	80.08
3	1	214.714	146.181	.587	-197.08	626.51
	2	172.000	170.152	.849	-307.32	651.32
	4	120.111	174.814	.958	-372.34	612.56
	5	-206.833	162.908	.711	-665.75	252.08
4	1	94.603	151.583	.971	-332.41	521.61
	2	51.889	174.814	.998	-440.56	544.34
	3	-120.111	174.814	.958	-612.56	372.34
	5	-326.944	167.772	.304	-799.56	145.67
5	1	421.548(*)	137.682	.027	33.70	809.40
	2	378.833	162.908	.152	-80.08	837.75
	3	206.833	162.908	.711	-252.08	665.75
	4	326.944	167.772	.304	-145.67	799.56

* The mean difference is significant at the .05 level.

Discussion

The purpose of this study was to determine if *Agreeableness* and its individual facets had a specific effect on team performance. What we have determined from our experiment is that there was not support for *Agreeableness* predicting team score, however there was support for *Agreeableness* predicting individual score. Our finding that *Agreeableness* predicts individual score and not team score is perplexing, but this may be due to the DDD task being designed to keep track of team scores, but inaccurately reporting individual scores. I am confident however that the team scores reported in the data are accurate as I witnessed some of the scores reported in the data being the administrator for several experiments. If there were any errors, I believe they took place in the reporting of the individual scores. Therefore the original hypothesis was not affected and the results showed that *Agreeableness* did not have a significant effect on team score. Another possibility for the lack of team *Agreeableness* level predicting team score may be due to communication only being allowed via the e-mail system and never occurring face-to-face during the task.

Not surprisingly, results were also non-significant for all of the hypotheses on the individual *Agreeableness* facets. It should be noted that the range of overall team *Agreeableness* was only between 2.3 and 3.783 with most scores falling around the middle of the scale (M=2.91). This range restriction prevented using a curvilinear regression and compelled us to use HLM in order to extract more information from the data. Perhaps using a ten-point Likert scale in future studies would allow for more variability and give more information about the personality aspects of the participants. In contrast to this, it may reproduce the results found in this study and provide more evidence for the lack of a curvilinear relationship between *Agreeableness* and team performance. Information on age, gender, and year in college were also used in the HLM analysis even though there were no specific hypotheses constructed for these variables and their effects can be found in Table 4.

Our hypothesis for GMA leading to better team performance was supported. This finding was consistent with past studies like Hollenbeck et al. (2002) and Neuman and Wright (1999) where intelligence had a significant effect on predicting team score. Our results may indicate that the task was intellectually demanding and this may have affected our results, serving to wash out any personality aspects. It may also mean that the effects of personality pale in comparison to how GMA affects team performance. Looking at the correlation between GMA and *Agreeableness* shows a slightly positive, but non significant relationship ($r=.113$, n.s).

Some limitations of this study could have been that the task was fairly complex and that while some of the participants seemed to pick up the task fairly easily, others clearly struggled. Better results might have been obtained by using a simpler task. Fatigue may have been a factor since the experiment took three and a half hours and the task lasted for 75 minutes. Apathy or lack of intrinsic motivation may have also played a part in the large number of participants who failed to score highly. It should also be noted that even given the chance many participants would not ask questions to the administrator or consult their quick reference guide for assistance. In addition, being unfamiliar with a computer may have hindered a participant's score. This is unlikely with the prevalence of computers in the classroom today, but still a possibility.

Training differences were not a factor as an ANOVA was performed on team score's and the experimenters who administered the task to them and it was not significant.

Future studies

For future studies, having more teams may be helpful to determine if there are really no specific effects for the facets of *Agreeableness.* In addition, specific hypotheses about general intelligence, experience with computer simulations, and other aspects of personality may be looked into and tested. Also, adding additional members to the team should lead to more variance amongst the *Agreeableness* within the team. This may lead to a clearer picture of why *Agreeableness* seems to affect individual performance, but not the overall team performance. Regarding the participants who failed to ask for help or search for aid in their quick reference guide, it may actually be more beneficial to have the quick reference guide present during training. All questions could be redirected towards the quick reference guide in order to prime more self reliance in the participants. However, this priming of self reliance may be detrimental to team performance due to a shift of focus to the individual. Perhaps a study using the two different training methods could be of interest. It would also be interesting to test teams who are currently working together in the work place and perhaps look at the most successful teams and work backwards.

In conclusion, this study was another important step in determining the relationship between an individual's personality and overall team performance. While we did not succeed in finding support for our mean *Agreeableness* hypotheses, we were able to test the individual facets of *Agreeableness* and their effects on team performance. Unfortunately, the hypothesis involving the individual facets were also not significant. We did find more support for higher levels of intelligence leading to better team performance. This may mean that intelligence is the one factor that must be present for team success.

References

Barcikowski, R. S.(1981). Statistical power with group means as the unit of analysis. *Journal of Educational Statistics, 63,* 267-285.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating Member Ability and Personality to Work-Team Processes and Team Effectiveness. *Journal of Applied Psychology, 83(3),* 377-391.

Bennet, R. P., & Carbonari, J. P. (1976). Personality patterns related to own-, joint-, and relative-gain maximizing behaviors. *Journal of Personality and Social Psychology, 34,* 1127-1134.

Bunderson, J. S. & Sutcliffe, K. M, (2003). Management Team Learning Orientation and Business Unit Performance. *Journal of Applied Psychology, 88(3),* 552-560.

Colquitt, J. A., Hollenbeck, J. R., Ilgen, D. R, Lepine, J. A., & Sheppard, L. (2002). Computer-Assisted Communication and Team Decision-Making Performance: The Moderating Effect of Openness to Experience. *Journal of Applied Psychology, 87(2),* 402-410.

Costa, P. T., Busch, C. M., Zonderman, A. B., & McCrae, R. R. (1986). Correlations of MMPI Factor Scales With Measures of the Five Factor Model of Personality. *Journal of Personality Assessment, 50(4),* 640-650.

Costa, P. T., & McCrae, R. R., (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4(1),* 5-13.

Ellis, A. P. J., Hollenbeck, J. R, Ilgen, D. R., Porter, C. O. L. H, West, B. J., & Moon, H (2003). Team Learning: Collectively Connecting the Dots. *Journal of Applied Psychology, 88(5)*, 821-835.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin, 121,* 149-167.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Heslin, R. (1964). Homogeneity of member personality and its effects on group problem-solving. *Journal of Abnormal and Social Psychology, 62*, 248-256.

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 expertise in hierarchical teams. *Academy of Management Journal, 41,* 269-282.

Hox, J. J. & Maas, C. J. M. (2002). Sample sizes for multilevel modeling. In: J.Blasius, J. Janis, I., & Mann, L. (1977). *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment.* New York: The Free Press.

Kilmann, R. H., & Thomas, K. W. (1975). Interpersonal conflict handling behavior as reflections of Jungian personality dimensions. *Psychological Reports, 37,* 971-980.

Klimoski, R. & Mohammed, S. (1994). Team Mental Model: Construct or Metaphor? *Journal of Management, 20(2),* 403-437.

Marks, M. A., (2000) A Critical Analysis of Computer Simulations for Conducting Team Research. *Small Group Research,* Vol. 31 (6), 653-675.

McCrae, R.R, & Costa, P. T. (1985) Updating Norman's "Adequate Taxonomy": Intelligence and Personality Dimensions in Natural Language and in Questionnaires. *Journal of Personality And Social Psychology. 49(3)*. 710-721.

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,* 9-45.

Neuman, G. A., & Wright, J. (1999). Team Effectiveness: Beyond Skills and Cognitive Ability. *Journal of Applied Psychology, 84(3),* 376-389.

Norman, W. T., (1963). Toward and adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66,* 574-583.

Porter, C. O. L. H., Hollenbeck, J. R., Ilgen, D. R., Ellis, A. P. J., West, B. J., and Moon, H. (2003) Backing Up Behaviors in Teams: The Role of Personality and Legitimacy of Need. *Journal of Applied Psychology, 88(3), 391-403.*

Raser, J. C., (1969). *Simulations and society: An exploration of scientific gaming.*

Rogelberg, S. (1999) Simulations for the study of teams. Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Schiflett, S. G., Elliott, L. R., Salas, E., and Coovert, M. D. 2004. *Scaled Worlds: Development, Validation, and Applications.* Abingdon: Ashgate Publishing Ltd.

Thompson, L. F. & Coovert, M. D. (2002) Stepping Up to the Challenge: A Critical Examination of Face-to-Face and Computer-Mediated Team Decision Making. *Group Dynamics: Theory, Research, and Practice, 6(1) 52-64.*

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (USAF ASD Tech. Rep. No. 61-97). San Antonio, TX: Lakeland AFB.

Williams, W. M., & Sternberg, R. J. (1988). Group intelligence: Why some groups are better than others. *Intelligence, 12,* 351-377.

Appendices

Appendix A

Personality Questionnaire

Session ID: ___________ A B

Listed below are phrases that describe people's behaviors. Please use the rating scale to describe how accurately each statement describes ***you***. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, and then circle the number to the right of the question.

		Very Inaccurate	Moderately Inaccurate	Neither Inaccurate nor Accurate	Moderately Accurate	Very Accurate
1	I trust others.	1	2	3	4	5
2	I believe that others have good intentions.	1	2	3	4	5
3	I trust what people say.	1	2	3	4	5
4	I believe that people are basically moral.	1	2	3	4	5
5	I get angry easily.	1	2	3	4	5
6	I like to solve complex problems.	1	2	3	4	5
7	I distrust people.	1	2	3	4	5
8	I suspect hidden motives in others.	1	2	3	4	5
9	I am wary of others.	1	2	3	4	5
10	I believe that people are essentially evil.	1	2	3	4	5
11	I would never cheat on my taxes.	1	2	3	4	5
12	I worry about things.	1	2	3	4	5
13	I fear for the worst.	1	2	3	4	5
14	I am afraid of many things.	1	2	3	4	5
15	I know how to get around the rules.	1	2	3	4	5
16	I cheat to get ahead.	1	2	3	4	5
17	I feel comfortable with myself	1	2	3	4	5
18	I prefer to stick with things that I know.	1	2	3	4	5
19	I can handle a lot of information	1	2	3	4	5
20	I obstruct others' plans.	1	2	3	4	5
21	I make people feel welcome.	1	2	3	4	5
22	I anticipate the needs of others.	1	2	3	4	5
23	I love to help others.	1	2	3	4	5
24	I love action.	1	2	3	4	5
25	I believe in human goodness.	1	2	3	4	5
26	I think that all will be well.	1	2	3	4	5
27	I act wild and crazy.	1	2	3	4	5
28	I am hard to get to know.	1	2	3	4	5
29	I have a lot of fun.	1	2	3	4	5
30	I take no time for others.	1	2	3	4	5
31	I would never go hang gliding or bungee jumping.	1	2	3	4	5
32	I go on binges	1	2	3	4	5

Appendix A (Continued)

		Very Inaccurate	Moderately Inaccurate	Neither Inaccurate nor Accurate	Moderately Accurate	Very Accurate
33	I hate to seem pushy.	1	2	3	4	5
34	I have a sharp tongue.	1	2	3	4	5
35	I contradict others.	1	2	3	4	5
36	I amuse my friends	1	2	3	4	5
37	I stick to the rules.	1	2	3	4	5
38	I use flattery to get ahead.	1	2	3	4	5
39	I use others for my own ends.	1	2	3	4	5
40	I hold a grudge.	1	2	3	4	5
41	I do not have a good imagination.	1	2	3	4	5
42	I talk to a lot of different people at parties	1	2	3	4	5
43	I laugh aloud.	1	2	3	4	5
44	I take advantage of others.	1	2	3	4	5
45	I believe that I am better than others.	1	2	3	4	5
46	I think highly of myself.	1	2	3	4	5
47	I have a high opinion of myself.	1	2	3	4	5
48	I seldom joke around	1	2	3	4	5
49	I find it difficult to approach others	1	2	3	4	5
50	I make myself the center of attention.	1	2	3	4	5
51	I like a leisurely lifestyle	1	2	3	4	5
52	I don't like crowded events.	1	2	3	4	5
53	I am concerned about others.	1	2	3	4	5
54	I have a good word for everyone.	1	2	3	4	5
55	I look down on others.	1	2	3	4	5
56	I take charge.	1	2	3	4	5
57	I love a good fight.	1	2	3	4	5
58	I yell at people.	1	2	3	4	5
59	I insult people.	1	2	3	4	5
60	I get back at others.	1	2	3	4	5
61	I like to tidy up.	1	2	3	4	5
62	I have a vivid imagination	1	2	3	4	5
63	I complete tasks successfully.	1	2	3	4	5
64	I excel in what I do.	1	2	3	4	5
65	I dislike talking about myself.	1	2	3	4	5
66	I consider myself an average person.	1	2	3	4	5
67	I seldom toot my own horn.	1	2	3	4	5
68	I am not interested in theoretical discussions	1	2	3	4	5
69	I like to visit new places.	1	2	3	4	5
70	I can manage many things at the same time	1	2	3	4	5
71	I have difficulty understanding abstract ideas.	1	2	3	4	5
72	I know the answers to many questions.	1	2	3	4	5
73	I boast about my virtues.	1	2	3	4	5

Appendix A (Continued)

		Very Inaccurate	Moderately Inaccurate	Neither Inaccurate nor Accurate	Moderately Accurate	Very Accurate
74	I avoid philosophical discussions	1	2	3	4	5
75	I am indifferent to the feelings of others.	1	2	3	4	5
76	I make people feel uncomfortable.	1	2	3	4	5
77	I turn my back on others.	1	2	3	4	5
78	I listen to my conscience.	1	2	3	4	5
79	I dislike being the center of attention.	1	2	3	4	5
80	I put people under pressure.	1	2	3	4	5
81	I pretend to be concerned for others.	1	2	3	4	5
82	I am not bothered by messy people.	1	2	3	4	5
83	I am easy to satisfy.	1	2	3	4	5
84	I can't stand confrontations.	1	2	3	4	5
85	I plunge into tasks with all my heart.	1	2	3	4	5
86	I seldom daydream.	1	2	3	4	5
87	I go straight for the goal.	1	2	3	4	5
88	I sympathize with the homeless.	1	2	3	4	5
89	I feel sympathy for those who are worse off than myself.	1	2	3	4	5
90	I feel others' emotions	1	2	3	4	5
91	I get others to do my duties	1	2	3	4	5
92	I value cooperation over competition.	1	2	3	4	5
93	I suffer from others' sorrows.	1	2	3	4	5
94	I am not interested in other people's	1	2	3	4	5
95	I tend to dislike softhearted people.	1	2	3	4	5
96	I do not enjoy going to art museums	1	2	3	4	5
97	I believe in an eye for an eye.	1	2	3	4	5
98	I try not to think about the needy.	1	2	3	4	5
99	I believe people should fend for themselves.	1	2	3	4	5
100	I can't stand weak people.	1	2	3	4	5

Appendix B

Background Information

The U.S. Military has been testing a new Unmanned Aerial Vehicles (UAV). During a snow storm the UAV was blown off course and contact was lost. It was determined that your team is the closest to the last known coordinates of the UAV. One team has already been sent out, but is now missing. You have been asked to help locate the lost team and recover the UAV. Good luck.

Antarctica is a continent covered with nearly 14 million square miles of ice. You and the other teams are at a politically neutral Antarctic Site, affectionately known as Station Blue, a research site specializing in geological activity and in the Earth's ozone layer. There are at most a few hours of daylight, the average daytime temperature is minus 13 degrees Fahrenheit, a 6-25 mph wind blows snow constantly, and visibility is usually less than 0.6 miles.

Station Blue is located 30 miles inland on an ice sheet at an altitude of 4,600 feet. Eighteen miles further inland, northeast of the station, is a mountain range with peaks as high as 11,000 feet. The terrain around the station is low undulating hills. To the east, there are canyons and bluffs formed by huge cracks and displacements of the ice sheet. The Antarctic wind has a deadly effect – in some places the wind can blow a sheet of snow and ice over crevices. Vehicles and individuals can be trapped or lost when they break through the sheet. In other places, the wind can form a natural bridge strong enough to support vehicles. Lastly, the wind can suddenly create blinding storms that can reduce visibility to zero and lower the temperature to minus 103 degrees Fahrenheit.

Station Blue has given your team permission to use two of its three snow cats. Specially designed to navigate Antarctica's terrain, the snow cat can carry your team members and a limited amount of supplies, has a top speed of 6 mph, but at its cruising speed of 4.8 mph they have a range of only 30 miles. Snow cats are equipped with communication and navigation equipment and a set of special sensors and probes. In addition, the snow cats are connected to geostationary satellites that can provide information about weather and geological events. Unfortunately, weather and terrain may interfere with satellite transmissions and disrupt radio communication with Station Blue.

You decide to use the snow cats to try to replicate the path of the lost team. The researchers have placed seismic monitors around the area – these monitors would have recorded the vibrations of a snow cat driving past it. There is a team at Station Blue to help access the supply depots and gas depots, and to provide guidance on the terrain. Since they have continual access to the satellite, they might be able to give you new information.

Appendix C

Tactical Information

STATION BLUE AND ITS RESOURCES

Seismic monitors: Seismic monitors usually examine the ice sheet for geologic activity, but they can indicate when a vehicle has passed and in which direction it was traveling. To obtain this information, you must process the seismic monitor with the correct level and combination of resources. You can apply more than the required resources but not less. If the monitor requires more supplies than you have, you can request another team to join you and combine resources to obtain the message. Seismic monitors take one to three steps to process. For example, three-step monitors need to be repaired, prepared, and then analyzed for information. Two-step monitors need to be prepared and analyzed. One-step monitors only need to be analyzed.

Waypoints: Certain seismic monitors yield key information, which needs to be read for the mission to be successful. Five or six seismic monitors in each scenario are designated as waypoints because they lie on the path taken by the fourth team. You will need to process these seismic monitors in order to uncover the path to the lost team. The closer a monitor is to the path of the "fourth" team, the higher the point value of a monitor. A monitor directly on the path of the "fourth" team will earn you 80 points; a monitor furthest away from the path of the fourth team will yield 10 points.

Processing the Monitors: Monitors require the following resources: mechanical, technical, and scouting. Mechanical ability refers to the materials needed to repair moving parts in a machine, like a vehicle or on a drilling machine. Technical resources enable the repair of circuitry in a computer or a digital computer chip. Scouting resources assist with the interpretation of the encrypted data on the monitors or at open locations. Resources can be depleted, but they can be replaced at specific locations. Some tasks may require more resources than you may presently possess. In those instances, you may have to request assistance from another team and pool your resources.

Satellite: The satellite provides relatively high-quality information and may indication the location of man-made objects or other geological clues to assist you. It also returns information about sectors where no clues were detected. Naturally, it provides weather and terrain information.

Clues: Clues provide information that can help you in your search and usually lay on the ice. You do not need to apply any resources to get a message from a clue.

TASKS

Terrain Tasks, Rough Terrain and Clearing Blocked Terrain. You are operating in a hostile environment that requires you to make decisions about which path to take. Hazardous terrain includes crevasses, mountains, and passes blocked by snow or avalanches. Hazardous terrain will slow your progress. Blocked terrain will stop your progress. Sometimes you will be able to clear the blockages; in other cases you must take a different path. Each vehicle has multiple units of terrain-clearing resources. Those resources are depleted with use but can be restocked at specific supply depots. Blue will receive information from the satellite concerning the location of the hazardous areas.

Repair tasks. You may encounter people who need mechanical help to repair equipment or machinery that has broken down. While these tasks will consume resources, they may yield useful information – e.g. "It's Cold Out Here" meaning you're searching in the wrong area.

Time-critical Emergency Tasks. Occasionally Station Blue will call you to render critical emergency assistance to another team. These are time-sensitive emergency tasks that may have life-threatening consequences if you do not help. If you don't respond to Station Blue's requests for help, your level of communication with Station Blue may be affected. <u>You will lose 100 points each if you do not attend to these emergency tasks within the allotted time</u>.

Non-critical medical tasks: Other medical requests may occur that are not critical or time sensitive, but your assistance may be rewarded with information. These tasks require medical personnel, which you may need anyway. I advise you to carry a medic for your own health.

Refueling: If your vehicle runs out of fuel everyone on your vehicle perishes! You can refuel at several fuel depots or via the movable fuel tanker. To do this you need to communicate with Blue and request refuel assistance. Your remaining fuel can be monitored.

Restocking: Your vehicle starts out with a finite number of resources. Resources are consumed by processing: seismic monitors, medical, repair, and emergency tasks. These resources can be replenished at one of several resource depots or at home base. Note that each REMOTE resource depot has resources to restock only one vehicle and is only accessible at certain times. The home base is always available for replenishing resources and has no limit on restocking capabilities. In order to replenish your supplies, you will need to communicate with Blue and request assistance.

SCORING

Score Display: You will be able to see your individual score and the scores of the other snow cats, if they are out there. BLUE will see the team score, which will be the sum of the scores from RED and PURPLE. Your score reflects four factors:

1. Points received after processing a seismic monitor,
    - The number of points you receive from processing a seismic monitor indicates how close you are to the lost party's path (10 points = far away, 80 points = very near). If you receive 0 points after processing a seismic monitor either someone has already processed it or you made an error in processing.
2. Attending to Emergency tasks
    - You earn 100 points if you attend to an Emergency task within the specified time; you lose 100 points if you don't.
3. Processing repair and medical tasks
    - If any task is processed by two snow cats simultaneously both players will receive the points for that task.
4. Recovering the UAV or the Lost Party.

Remember that to receive points on a task requiring coordinated efforts (e.g. 3 medical units and 3 technician units), you must have them both process the task simultaneously. If, for instance, the medic finishes processing the task that requires the use of both a medic and a technician before the technician is instructed to process the task, then the task will disappear and you will receive no points. All necessary personnel must be begin processing the task before the first person finishes processing or you will not be given credit for completing the task.

Point Allocation:

300 points:	Recover the UAV.
300 points:	Recovering the lost team.
100 points:	Render emergency assistance
-100 points:	**Failure to responded to emergency task within the allotted time**
50 points:	Assist with repair or medical requests
10 - 80 points:	Process seismic monitors (high points = close to lost party's path)

Important points to remember:

- Make sure you attend to the emergencies within the time allotted!
- It is important to maintain fuel levels and to refuel when necessary. Your snow cat will be immobile if fuel drops below 150 pounds. If your vehicle runs out of

fuel, everyone on your vehicle will perish and you will be out of the game. You must also maintain sufficient personnel resources on your snow cat to be able to process the seismic monitors and other tasks.

- Remember to load personnel onto your snow cat, as this will allow them to travel faster**.**
- Apply at least the required amount of resources to complete the seismic monitor and medical/repair tasks. Remember, if the task requires more expertise than you have on board your cat, you can request another team to help by combining their resources with yours.
- *Do not forget to fill up at the supply depots if you run low on resources (i.e. medic units, technician units, mechanic units, etc.)*