Additional Evidence:

Risk of Performance Decrements Due to Inadequate Cooperation, Coordination, Communication, and Psychosocial Adaptation within a Team

Emily M. David, Cristina Rubino, Kathryn E. Keeton, Christopher A. Miller, Holly N. Patterson, ***An Examination of Cross-Cultural Interactions aboard the International Space Station (ISS).*** National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, Texas September 2010.

Veronica Maidel, Jeffrey M. Stanton, ***Unobtrusive Monitoring of Spaceflight Team Functioning: Literature Review and Operational Assessment for NASA Behavioral Health and Performance Element***. School of Information Studies, Syracuse University, August 2010.

Jacquelyn Ford Morie, Gustav Verhulsdonck, Rita Lauria, ***Operational Assessment Recommendations: Current Potential and Advanced Research Directions for Virtual Worlds as Long Duration Space Flight Countermeasures***. yle Laboratories and Institute for Creative Technologies, University of Southern California.

David Musson, ***Final Report: Investigating the Influence of Personality on Astronaut Career Performance***. Prepared by Contractor for NASA, September 2010.

Raymond A. Noe, Ali McConnell Dachner, Brian Saxton, ***Team Training for Long Duration Missions in Isolated and Confined Environments: A Literature Review, Operational Assessment & Recommendations for Practice and Research***. Prepared for NASA by Department of Management and Human Resources, Fisher College of Business, The Ohio State University September 2010.

Human Research Program
Behavioral Health and Performance Element

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas

An Examination of Cross-Cultural Interactions aboard the International Space Station (ISS)

Human Research Program
Behavioral Health & Performance Element
Space Medicine Division

September 24, 2010

Emily M. David, M.A.

Cristina Rubino, M.A.

Kathryn E. Keeton, Ph.D.

Christopher A. Miller, Ph.D.

Holly N. Patterson

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas

Table of Contents

NASA HUMAN RESEARCH PROGRAM

BEHAVIORAL HEALTH AND PERFORMANCE ELEMENT

An Examination of Cross-Cultural Interactions aboard the International Space Station (ISS)

Introduction

The Behavioral Health & Performance Element (BHP) is one of six elements within the Human Research Program and is comprised of four Risks, namely the Risk of Behavioral Conditions (BMed), the Risk of Psychiatric Disorders (BMed), the Risk of Performance Decrements due to Inadequate Cooperation, Coordination, Communication, and Psychosocial Adaptation within a Team (Team), and the Risk of Performance Errors due to Sleep Loss, Circadian Desynchronization, Fatigue, and Work Overload (Sleep). BHP is tasked with designing, implementing, and managing research tasks that will develop tools, technologies, countermeasures, and other mitigation strategies to help support the crew on long-duration missions.

Within the Team Risk, seven specific gaps were identified in which critical knowledge is unknown or an adequate mitigation strategy has not yet been developed. Team Gap 4 poses the question "Given the context of long-duration missions, what are the optimal ways to select individuals and compose crews to ensure/optimize/facilitate task performance, teamwork, and psychosocial performance?" Additionally, Team Gap 5 asks, "Given the context of long-duration missions, what are the optimal ways to train crews, leaders, and ground support to ensure/optimize/facilitate task performance, teamwork, and psychosocial performance?" To address these gaps, it is important to first have a clear understanding of the individual differences contributing to astronaut success and well being (i.e., task performance, teamwork, and psychosocial performance). Accordingly, by investigating whether cultural issues impact communication and team interaction, researchers will be well positioned to address the selection, composition, and training questions outlined in Team Gaps 4 and 5.

Investigating the impact of culture within the astronaut population is a timely issue given that problems related to human interaction (including those that are culturally based) are likely to increase in prevalence as we move toward extreme long-duration space flight. In early space flight, the short duration of the missions caused psychological and interpersonal issues to be relatively inconsequential. Although one can cope and even ignore interpersonal conflicts and communication difficulties in the context of a one- or two-week mission, these factors can become a chronic stressor on long-duration missions given the heightened isolation and greater amount of down time that crewmembers will experience. Further, it is likely that successfully completing a long-duration mission will require the cooperation of multiple nations and individuals with a variety of cultural backgrounds. Given these projections, the main objectives of this project were to: (a) identify, document, and describe any existing issues in communication occurring in cross-cultural teams and (b) examine whether cultural differences in behavioral outcomes exist.

Background

In response to increased globalization and the advancement of communication technologies, researchers have increasingly studied the impact that national culture can have on interpersonal communication, relationships, and work outcomes (e.g., Davison, 1994; Kanter & Corn, 1994; Snow, Snell, Davison, & Hambrick, 1996). Although some researchers have noted a number of positive outcomes resulting from culturally diverse teams such as varied perspectives and skills (e.g., Maznevski, 1994; Watson, Kumar, & Michaelson, 1993), others have noted that communication decrements can hinder team effectiveness and cohesiveness (Bantz, 1993; Matveev & Nelson, 2004). Because multicultural team members vary in their environmental perceptions, motives, and communication norms, several misunderstandings can occur that may impede effective team functioning. Indeed, research has shown that when a group consists of members from cultures with varying levels of power distance (Hofstede, 1980), for example, communication and leadership difficulties often arise (Bantz, 1993). Other studies have revealed that poor cross-cultural communication can lead to lowered social cohesion (Shaw, 1981), increased conflict, and lowered performance (Shenkar & Zeira, 1992), and social stigmatization (Molinsky, 2005).

Hofstede's (1980) cultural dimensions, which provide one of the most well-established methods of classifying culture, theoretically underlie many of these issues (although other researchers have proposed alternative dimensions to compare multiple cultures, e.g., McSweeney, 2002). Four dimensions (i.e., individualism/collectivism, power distance, uncertainty avoidance, and masculinity/femininity) were produced as an outcome of the original study that spanned over 117,000 employees from 72 different countries. More recently, Hofstede (2001) added a fifth dimension: long- versus short-term orientation. The existence of these dimensions was replicated, and the dimensions were found to be stable across time (Hofstede, 1980). The first, individualism versus collectivism, is arguably the most widely cited and reflects the degree to which individuals are integrated into groups (Hofstede & McCrae, 2004). Individualism involves a tendency to be more interested in one's own goals, needs, and interests, while collectivism involves a tendency to be more concerned with the needs, interests, and goals of the group or society to which one belongs (Vodosek, 2009). The second dimension, power distance, is an indicator of the equality or inequality of power reinforced by a society and refers to the extent to which hierarchies are emphasized and powerful individuals are treated differently. Uncertainty avoidance, the third dimension, is defined as the extent to which members of a particular culture are threatened by ambiguous situations. Masculinity versus femininity, the fourth dimension, refers to either a focus on material objects and success (i.e., high masculinity) or caring for others and quality of life (i.e., high femininity). This dimension also highlights cultural attitudes towards the roles of men and women (Hofstede, 1980; 2001). Finally, long- versus short-term orientation refers to the degree to which members of a particular culture are encouraged to delay gratification of material, social, and emotional needs.

Whereas each of these culture facets have been shown to impact the way that members of a particular culture think and behave, we contend that power distance and individualism-collectivism are particularly applicable to cross-cultural communication in the space flight context. Indeed, the individualism-collectivism and power distance dimensions have been related to both psychological and work outcomes (Fiske, Markus, Kitayama, & Nisbett, 1998). For example, Hofstede (1980) stated that management style is influenced by a combination of these two facets, and Camiah and Hollinshead (2003) found that these cultural dimensions greatly impacted communication difficulties between Russian and American managers. Additionally, individualism-collectivism and power distance also may influence verbal behavior norms, as well as the perception of these norms, which vary to a significant degree between cultures (e.g., Kowner & Wiseman, 2003). For example, several studies have found that individuals from more collectivistic cultures experience more communication apprehension than do North Americans (Watson, Monroe, & Atterstrom, 1984).

Culture and NASA

The U.S. space program has seen a steady evolution in NASA missions, each with a different aim and at different levels of international cooperation. The first American human space program was the Mercury program, which was tasked with investigating how to reach, live in, and return from space (Smith, 2000). Following this effort, the Gemini program devised a way to dock with other spacecrafts in orbit and the Apollo program allowed astronauts to explore the lunar surface (Naval Historical Center, 2003). The missions within these three programs were characterized by a relatively short duration (approximately one week) and devoid of international partnerships (Smith, 2000).

Starting in 1975 with the Apollo-Soyuz test project, American astronauts and Russians formed a series of international partnerships that changed the homogenous face of crews. In addition, the length of missions shifted to be longer in duration (i.e., several months at a time in space) in lower-Earth orbit beginning with Skylab in 1973. International cooperation and longer missions were then combined with the introduction of the Shuttle-Mir program and finally with the birth of the International Space Station (ISS), which began construction in October of 2000 (Oberg, 2007). Although other international partners are sometimes part of the crew, the primary cultural contrast aboard ISS is the juxtaposition of Russians and Americans. This is because all ISS crews have historically had at least one Russian and one American crew member, and both countries jointly plan missions (Boyd, Kanas, Gushin, & Saylor, 2007). Additionally, there are separate Russian and American segments, and both countries' science experiments are given equal priorities (Boyd, 2005). In line with the preceding information, we chose to limit our focus to these two cultures in the present studies.

As Russians and Americans are dissimilar along the individualism-collectivism and power distance dimensions, crew interaction aboard the ISS, which consists primarily of Russian and American astronauts[1], may be influenced by existing cultural differences. Although Russia was not included in Hofstede's original research, researchers have since conducted studies identifying this nation's standing on the various dimensions (e.g., Bollinger, 1994). For example, Russians tend to be more collectivistic in orientation and place a greater emphasis on group harmony, cooperation, and relationships. Americans, alternatively, are more individualistic and place a high value on autonomy, task performance, and self-

[1] Although current ISS practice involve a small proportion of crew members from other cultures and nationalities, it seemed productive to begin our research with a focus on Russian and American interactions.

interests. Additionally, in contrast with Americans, Russians are higher in power distance and make a greater distinction between people of high and low power and status. This can become particularly problematic in multicultural groups, where low-power distance individuals tend to treat everyone the same whereas high-power distance people make greater distinctions in the way they communicate with individuals in different social or organizational positions (Bantz, 1993). Accordingly, Russian culture is characterized by high power distance and a collectivistic orientation, which translates into a very autocratic management style (Bollinger, 1994). Low-power distance and individualistic Americans, alternatively, expect much more autonomy at work as well as ample rewards for individual performance.

A limited set of quantitative and theoretical articles have been published specifically focusing on cross-cultural issues that occur on longer duration space missions. For example, a review piece by Boyd (2005) discusses relevant cultural differences among crew members in values, emotional expressivity, personal space norms, personality, foreign language competence, gender norms, personal relationships with co-workers, and cultural heterogeneity. She highlights that, if left unmanaged, these disparities have the potential to impact not only individual well being but also mission success (Boyd, 2005). In a later review, Kanas and colleagues pointed to differences in emotional expressivity, manifestations of depression, cognitive and decision making styles, norms of hygiene and privacy, and social behavior norms surrounding guests and meals (Kanas, et al., 2009).

Sandal and Manzey (2009) surveyed members of the European Space Agency (ESA) who reported that the greatest challenges of working on a multicultural team are differences in leadership preferences and conflicting management styles. This survey also revealed that a key challenge listed by the ESA members was interacting efficiently with other crew members (Sandal & Manzey, 2009). This finding further highlights the need to understand the impact of differing cultural backgrounds on communication in the space flight context. Researchers also have identified other outcomes of cultural differences on the space station. Specifically, Boyd, Kanas, Gushin, and Saylor (2007) found that cultural background predicted differing manifestations of distress. For astronauts aboard ISS, anxiety was a primary factor associated with depression, whereas for cosmonauts, fatigue was a primary indicator of depression.

Study Overviews

In Study 1, we outline case studies of instances aboard the ISS where we focus on the impact of power distance and individualism-collectivism (and their impact on direct and indirect communication), general cultural knowledge, and high-and low-context cultures (Hall & Hall, 1990). In contrast to the United States, Russia is considered a high-context culture (Brett, Adair, Lempereur, Okumura, Shikhirev, Tinsley, & Lytle, 1998). This is often displayed in the use of a circular manner of communication (Berdiaev, 1990), which people from low-context cultures may construe as secretive or excluding of those around them (Rajan & Graham, 1991; Lewis, 1996). High-context culture is also displayed through a host of commonly accepted nonverbal behaviors. For example, in Russia, public physical contact including hugs, backslapping, and kisses on the cheeks are common among friends or acquaintances and between members of the same sex (Morrison, Conaway, & Borden, 1994). Additionally, Russians tend to leave less interpersonal distance when conversing. It is also more common to hold a gaze and touch the other person when communicating in Russia than in America (Watson, 1970).

Conversely, in low-context cultures such as the United States, the message one is attempting to convey is stated explicitly. Alternatively stated, the actual words expressed are a literal transcription of the messenger’s beliefs, attitudes, and requests (Schwartz, 1994). Individuals from low-context cultures also often have very visible external reactions when communicating, and tend to be much more task focused than relationship focused. In high-context cultures such as Russia, communication norms dictate that one should not be abrupt and should instead rely on tone, implications of words, and the nonverbal context to convey one's true meaning. In addition, there is heavy use of covert messages and metaphors, and much of the meaning relies on contextual and background knowledge (Morrison et al., 1994). In high-context cultures, the prime objective of communicating is preserving relationships and saving face rather than efficiently conveying task objectives (Matveev & Nelson, 2004).

Cross-cultural research also has suggested that culture impacts the way we perceive social behavior and the attributions made about that behavior (Pekerti, 2005). For example, people from individualistic cultures tend to make dispositional or internal attributions about social events rather than looking to contextual explanations. Those from more collectivistic cultures, in contrast, tend to attribute events to fate or other contextual factors (Betancourt & Weiner, 1982; Miller, 1984; Morris & Peng, 1994; Schuster, Forsterlung, & Weiner, 1989).

For the second study, we combined motivational facets derived from individualistic/collectivistic and high/low context cultures. Specifically, individuals who use communication to build and maintain relationships (i.e., collectivistic orientation) often do so by adopting an indirect communication style and are called sociocentric. Alternatively, those individuals more concerned with efficiently accomplishing task objectives (i.e., individualistic orientation) often adopt a more direct, no frills communication style and are called idiocentric (Triandis, McCusker, Betancourt, Iwao, Leung, Salazar et al., 1993). In study 2, we looked at whether or not American and Russian crewmembers differed in communication style along these two dimensions. Specifically, we expected that Americans would demonstrate more behaviors indicative of an idiocentric communication style, whereas Russians would demonstrate more behaviors in line with a sociocentric communication style.

Another cultural issue impacting communication is the idea that politeness perceptions vary from nation to nation. Although attempts at politeness are perhaps more closely seen as aimed at making people feel at ease and even flattered, sociolinguistic theory (Brown and Levinson, 1987) documents that politeness is how power and familiarity (including team) relationships are negotiated and maintained, and how imposition, urgency, and obligation are established. Although politeness is seen as serving these roles in all cultures, the specific methods by which politeness is conveyed frequently differ across cultures—which establish opportunities for mismatches and misinterpretations. Brown and Levinson (1987) devised a model of politeness to capture how these discrepancies come into play. Although they contend that the idea of politeness is universal, cultures may prefer negative politeness (e.g., deference, being indirect, "if you don't mind") while others prefer positive politeness (e.g., establish familiarity with person, informal, jokes; Brown & Levinson, 1987). Also, not all strategies (e.g., a thumbs-up gesture) are perceived equivalently by all cultures. These politeness or redressive strategies are used to mitigate any potential face threat caused by a request or other action. If the appropriate redressive strategy is not chosen to counter the potential face threat, imbalance may occur that causes the speaker to be perceived as rude or even hostile. Because the cooperation aspect of teamwork involves a shared perspective and politeness for people to comply with requests, differing notions of politeness can certainly affect team functioning. For this reason, in Study 3 we examined the various components of face and politeness to examine imbalances that have occurred aboard the ISS.

Below we describe the method and results for the three studies that utilized videos taken from aboard the ISS, which captured interactions between American astronauts and Russian cosmonauts. Specifically, these investigations focused on exploring how various cultural facets, politeness norms, and communication style differences are exemplified in interactions aboard the ISS.

Study 1

Objective

The objective of Study 1 was simply to determine whether we could find cultural communication disparities in the video records available in the Video Assessment Management System (VAMS). VAMS is an online film repository that includes all video that has been collected by NASA including those videos taken from the early space programs (including Apollo) to the present videos that are downlinked from Shuttle and ISS.

Method

The videos selected for this first study were not a random sample from the whole collection of videos. Instead, we selected five video clips taken aboard the ISS from the online film repository, VAMS. These clips met the following criteria: a) clear audio to understand what was being said, b) at least one person each from Russia and the United States communicating with one another, and c) perceived cultural disparities observed during the communication..

A case study analysis approach was then taken to examine cultural differences observed in the videos. Below, each video is summarized and subsequently analyzed to highlight how these differences are manifested in crew member interactions.

Case Study Analysis

Videos 1 and 2 – Power Distance and Group Participation

The first two videos demonstrated communication difficulties between Russian cosmonauts and American astronauts resulting from differences in power distance. As stated previously, Russians are higher in power distance than Americans (Bollinger, 1994). In the first video, DVCAM 32 (taken during an ISS Expedition), some members of the Expedition crew are interacting with members of the STS crew. The group consists of five Americans, two

Russians, and an Italian. The camera shows the two Russian cosmonauts, who are part of two different crews, sitting together in the corner. The topic being discussed is how tired the STS crew feels, which leads to a story about how they accidentally fell asleep an hour early the day before. Here is a sample of some of the dialogue:

All: [Laughter]

Italian: "…wake up earlier! Sleep well."

American 1: "Oh I will. I'm almost asleep right here just standing here. And then you guys…"

Various: "yeah." "Yeah!"

American 2: "Why are *you* tired?"

American 3: "I don't know what did we do today?"

American 2: "We went to bed an hour early!"

American 3: "We didn't know about it though!"

American 1: "Yeah we went to sleep and then mission control called in…"

American 3: "Are you guys on the same schedule as us?"

American 4: "We are about a half-hour off."

The conversation is animated and details are supplied in a choppy manner from a number of different speakers around the room. Questions are thrown forth by various participants. Although the conversation is loud and informal and people are interrupting each other frequently, the two cosmonauts in the corner stay distinctly quiet throughout the whole exchange.

In the second video, DVCAM 05 (taken during an ISS Expedition), we once again see the same two crews interacting in a social setting. This time, they are having a makeshift luau with everyone dressed in Hawaiian shirts eating shrimp and pork chops. There is a lot of lighthearted conversation going on as the crew members work to distribute the 'treat' food items and get into their shirts and sunglasses. There are even jokes about one of the American's hair and then another astronaut calls out "Get Steve!" A life-sized cutout of Steve Martin is then produced and invited to join the party. There are once again several Americans shouting back and forth with stories and jokes, but the Russian crew members are noticeably quiet and once again sitting in close proximity to one another. There are even several times that an astronaut attempts to include one of the Russian crewmembers in the conversation, but the cosmonaut limits his answers to one line or less without follow up. The only time one of the Russian crew members participates in the conversation at length is when several of the members go away and the group is reduced to his three crew members (and he is commander).

The disparity in participation observed in these two videos can be explained by differing power distance norms in the two countries. In low-power distance cultures, no matter one's social or hierarchical standing in an organization, everyone is encouraged to speak up. Indeed, having the courage to speak to one's superior is seen as a virtue (Kirby & Barger, 2009). In high-power distance cultures, however, such differences in power are seen as instrumental in keeping peace and harmony among groups. In large groups, then, the average member from a high-power distance culture may not be inclined to speak up in group conversations. Such actions may be seen as trying to stand out or not knowing one's place. Indeed, research suggests that members of collectivistic cultures experience more communication apprehension than those from individualistic cultures (Watson et al., 1989). Accordingly, the Russian cosmonauts may have refrained from group conversation, particularly when in the presence of commanders or people with more experience, to show deference and respect (Kirby & Barger, 2009).

Videos 3 and 4 – Individualism – Collectivism and Direct Communication

The third and fourth clips are taken another Expedition crew, which is comprised of two Russian cosmonauts and one American astronaut. In the first clip, DVCAM 38 (taken during an ISS Expedition), two cosmonauts are getting ready for an EVA while an American astronaut is videotaping them. The American filming the EVA preparation asks the two cosmonauts questions as well as tells them to smile at the camera. When a Russian cosmonaut complies with the American astronaut's request, the astronaut states "good job." The following is an excerpt taken from this video.

American: "Smile for the camera. This'll go down today. Thumbs up."

Russian: [Does not seem to respond as he is busy putting on his suit].

American: "Ready?"

Russian: "Ready."

American: "Everything is good?"

Russian: [Gives a thumbs up].

American: "Good job."

American: "(Calls his name). Smile. Ready to go?"

American: "(Calls his name)."

America: "(Calls his name)."

Russian: [nods]

Russian: "Ready."

American: "Good job."

In the second video clip, DVCAM 42 (taken during an ISS Expedition), the same cosmonaut is about to leave for the spacewalk as the American astronaut continues to film. The following excerpt is taken from this video clip.

American: "Very good."

American: "First time that hatch is open in space."

American: “Good job. Good pilot. Still works.”

Russian: “See it.”

American: “Let’s go look.”

American: “Here, film it. Take a picture.”

Russian: “I am the first touch it.”

American: “Show me.”

American: “Very good. First time.”

American: “Here, take one this way. Good job.”

These clips highlight cultural differences in direct versus indirect communication styles as well as in giving individual recognition, which are associated with Hofstede’s individualism-collectivism cultural dimension. First, individualistic cultures tend to communicate in a direct manner, which explicitly communicates the wants, needs, and desires of the speaker (Jandt, 2006). In collectivist cultures, the indirect communication style is preferred. With indirect communication, “the wants, needs, and goals of the speaker are not obvious in the spoken message” (Jandt, 2006, p. 162). In the first clip, the American astronaut is much more direct as he calls the cosmonaut’s name repeatedly even when the cosmonaut does not give a response. Also, in both clips, the American directly requests that the cosmonaut smile or do a ‘thumbs up’ without stating “please” or “would you mind.” Although this may not be perceived as an order in cultures that are accustomed to communicating in a direct manner (i.e., individualistic cultures), Russians, who come from a collectivist culture and are therefore accustomed to more indirect forms of communication, may perceive this as an order rather than a request. Although the Russians in these clips do not respond negatively, there is a notable difference in their communication style. The Russians in both clips do not tend to respond directly or make similar requests of the

American. Instead, they continue to focus on their work tasks and respond to questions as needed.

Another cultural difference that is exemplified in these clips is the differing importance ascribed to recognizing contributions on an individual level. Individualistic cultures are more prone to recognize and reward individual behavior. However, in collectivist cultures, individuals are more concerned with the success of the group and therefore do not seek individual recognition. Although the American is frequently positively praising the Russians in both clips (e.g., “good job, good pilot”), the Russians may feel uncomfortable being singled out as they are not used to receiving direct praise for their individual accomplishments.

Video 5 – Culturally Ingrained Knowledge

The fifth video, DVCAM 05 (taken during an ISS Expedition), shows the arrival of the STS crew (five Americans and one Russian crew member) on the space station. They are greeted by the Expedition crew, which contained two Russian and one American crew member. Despite the greater number of American crew members, most of the greetings and conversations were conducted in Russian. A translator was contracted to translate this dialogue and dialogue from other video clips as well as to give an idea of the context behind any of the colloquial expressions used. Below is an excerpt of the conversation that occurred as the hatch was opened and the STS crew entered the space station (note that this entire conversation occurred in Russian):

Russian 1: “True...true...presents.”

Russian 2: “Good ones…with arrival!!” (*Transcriber’s note: “with arrival” is a Russian greeting to guests with a similar meaning as “welcome!”*)

Russian 1: “Hi!”

Russian 3: “Congratulations! Congratulations!”

Russian 2: "We spoke with [Nickname] (*Transcriber's note: short for [longer Russian name]*) and I congratulate you!"

Russian 3: "His birthday!" (*Transcriber's note: Russian word for word translation of "happy birthday!" is "I congratulate you with birthday!"*)

American 1: "Congratulations!"

Russian 1: "[American astronaut name]!! Good to see you! I must kiss you. Three…" (*Transcriber's note: Russian tradition is to give three kisses on the cheeks upon meeting someone*)

Later, after a safety briefing in English:

Russian 2: "We are going to capture this historic moment."

Russian 2: "If *you* have no objections, of course." (*Transcriber's note: The word "you" that is used here is the formal form of addressing another person in Russian. This form of address is usually used with strangers, with those who are much older, or those in a position of authority. Here it is used as a joke*).

As can be seen here, there are a number of cultural nuances that must be understood for cross-cultural communication to be effective. Even a thorough knowledge of the Russian language might not fully prepare the American crew member for the cultural implications of what is actually being said if one was only focused on literal translations. For example, without the proper cultural knowledge, Americans may misunderstand "with arrival" as the first part of a sentence, and may be left waiting to hear what has arrived instead of understanding this to mean "welcome." This tentative and emotionless (or even confused) reaction may in turn be perceived as rude by the Russian greeter, leading to an early communication breakdown. Similarly, one might misunderstand "I congratulate you" for a job well done while docking or completing some other task objective, rather than wishing someone a happy birthday. Even in this conversation, there is some evidence of misunderstanding, as one of the cosmonauts must clarify the reason for the congratulatory message by saying "his birthday." In addition, knowing that three kisses are traditional would help to avoid any awkwardness associated with one person coming in and the other pulling away. Once again, we see clarification being used as the cosmonaut tells the American crewmember how many kisses to expect (i.e., "three"). Knowledge of this custom may also help to avoid any misunderstandings regarding why the cosmonaut feels he must kiss the American (female) astronaut. As kissing is not a customary greeting among friends in the United States, such traditions should be explained and understood to avoid miscommunication.

Finally, the grammatical distinctions between things like the formal and informal "you" reflect a subtle distinction that can be used to give context and meaning to what is actually being said. As discussed previously, Russia is a high-context society (Brett et al., 1998). Therefore, Russians tend to rely on the tone and context of what is being said to carry much of the true meaning. In this example, the crew member is using a very formal version of the word "you" when requesting to take pictures. Although the content of the words literally are asking something of someone else with a high level of deference and formality, in this instance, the opposite meaning is actually intended. Specifically, in the context of their apparently close relationship and following laughter and hugs, using such a formal "you" is an instance of sarcasm. Such a joke has the intention of showing the closeness between the two individuals rather than showing any sort of stilted formality as the words would suggest. This is important to understand from an American viewpoint, because otherwise one might mistakenly understand the conversation to have a completely different implication.

Conclusions & Implications

The results of these case studies clearly demonstrate that cultural differences do indeed impact communication between astronauts aboard ISS on, at least, some occasions.

For example, these clips show instances where the high-power distance and collectivistic Russian crew members are apprehensive about participating in large group conversations. Further, they highlight differences in direct and indirect communication as well as individual recognition preferences associated with individualistic and collectivist cultures. Finally, we noted several instances where specific knowledge of Russian culture is needed to fully understand what is being communicated.

Certain limitations to this approach must be addressed. First, we purposely selected the videos that we perceived as displaying difficulties in cross-cultural communication. Although these clips are illustrative of cross-cultural communication challenges that have occurred aboard ISS, they are not a random sample from the whole collection of videos. It is important to keep in mind, however, that the videos collected in this repository were previously filtered by both the crew (as they have discretion to turn the video cameras on and off) and possibly by NASA officials on the ground as well. Accordingly, the fact that we were able to identify any evidence of such instances where culture may have impacted communication warrants future research to further explore this topic. A second limitation exists within alternate explanations to the various communication phenomena observed in the videos. For example, we described the first two videos within the context of power distance differences. An alternative explanation for the non-participation of the Russians in the larger group was that both conversations were conducted in English and there may have been a language barrier preventing their full immersion in the conversation. Even if this were the case, however, language barriers still represent communication difficulties with cultural relevance.

In conclusion, we believe these case studies highlight the fact that cross-cultural communication issues do currently occur aboard the ISS. The prevalence and intensity of the resulting misunderstandings is something that future research should explore. These issues are likely to increase in importance as more and more international partners send their representatives into space. Instead of learning cultural nuances for only one other nation, present and future astronauts will have to learn a variety of cultural norms from Japan, Canada, and European nations. Future research also should be directed at identifying potential selection criteria and training modules aimed at increasing cross-cultural communication competency.

Study 2

Objective

The ultimate goal of this case study was to investigate whether or not significant communication differences exist in multicultural dyads when compared to same culture dyads—that is, to establish whether ISS residents interact differently with members of their same culture than with those from a different culture. Additionally, we hoped to observe any cultural differences in influence outcomes observed in the videos.

Method

This study utilized a modified version of the culture by condition comparison method outlined by Pekerti and Thomas (2003). Similar to the first case study, we again utilized the VAMS online repository system to select the video clips that would be coded. We first identified 20 video clips from the same VAMs site used in the previous study that were a) 10 minutes or less in duration and b) featured verbal and nonverbal interactions among the astronauts. Ten of the videos had a Russian and American astronaut interacting and the other 10 had two American astronauts interacting. We then compiled these file names into a spreadsheet and labeled individuals in the videos as either Astronaut A or Astronaut B (e.g., Astronaut A has a blue suit and a mustache) to reduce bias in the coding.

Three independent raters were then given the materials and a calibration training session that involved going over the operational definitions for each of the rating categories and then going through a sample video and coding it together to get an idea of what different verbal and nonverbal behaviors may look like. Raters were instructed that a single frequency of a behavior should be defined as a sentence or less in duration, and that only interactions between the two focal astronauts should be counted. The raters then coded frequency counts of behaviors for each of the 20 videos.

In Appendix A, we include the full behavioral indicators as listed in the survey used by the raters to make the frequency counts (Pekerti & Thomas, 2003). Below we list the operational definitions provided to the raters. The first eight behaviors reflect idiocentric tendencies whereas the next four reflect sociocentric tendencies. The final two represent the behavioral outcomes that we measured:

- **Expressive -** outwardly displaying emotions in pitch of voice, laughter, anger, and also not being extremely "technical" or strict/formal in communication style; outwardly displaying emotions through hand gestures, facial expressions, smiling and loud volume, etc.
- **Dominant -** overpowers others, projects/talks louder than other people they are speaking with, talks more, takes the reins of a task, etc.
- **Initiate Action -** begins the conversation, offers assistance more than conversing partner, positions body to be the center of attention, etc.
- **Aggressive -** sound of voice has aggressive tone, seems brash, asserts more power in actions than others involved, etc.
- **Logical and Systematic -** presents well thought out ideas, convincing, following procedures, etc.
- **Regulates flow -** communicates well with everyone, fills in conversation gaps when others fall silent, keeps conversations going, changing topics, etc.
- **Concerned with Finishing the Task –** comments to steer the conversation back towards the task at hand and away from leisure or tangential topics; body and verbal language imply one is unconcerned about any other topics than getting the job done, etc.
- **Strong Opinions -** speaks openly about thoughts/opinions; not afraid to say how they really feel about something, especially if it goes against popular belief, etc.

- **Agreeable -** doesn't show much self-confidence in opinions or thoughts, doesn't share opinions, goes with majority or other CMs ideas/thoughts/actions, etc.
- **Avoid arguments** – doesn't go against another's ideas or thoughts, changes topics when someone appears aggressive, etc.
- **Shift opinions -** readily changes train of thought or idea to coordinate with others, etc.
- **Eye contact -** physically making eye contact with other crew members, etc.
- **Influence** – does this person influence (as indicated by facial/body/language expression) another person to change their way of thinking or action, etc.
- **Change behavior** – does this person engage in different verbal or nonverbal reactions as a result of the words or actions of the other person, etc.

Results

Once the ratings were completed, we first assessed the degree of interrater reliability to justify aggregation. According to Landis and Koch (1977), an interrater agreement of .41-.60 is considered moderate, a range of .61-.80 is considered substantial, and .81-1 is considered almost perfect agreement. Given these criteria and, as shown in Table 1, all of our videos had adequate levels of agreement to justify averaging the ratings of the three raters before continuing our analyses. The average of all 20 intraclass correlation ICC values was .59 with a standard deviation of .14. Additionally, all of the ICC values were significant at the .05 level.

Accordingly, we averaged the ratings for each category across the three raters. We then conducted T-test analyses to determine whether there were significant differences in the communication styles used by individuals from certain cultures and also in various dyad composition types (e.g., do Russians use a more sociocentric communication style when interacting with Americans?). We also looked to see if there were group differences in the two outcome variables (i.e., influencing another crew member or changing one’s own behavior).

Table 2 summarizes the results of these analyses. As shown here, there were no significant differences between the crew members in the videos where two American astronauts were interacting. These results suggest that with regards to idiocentric, sociocentric, and outcome behaviors, Americans interacting with one another were roughly equivalent. In the videos where a Russian and American interacted, we expected to see significant differences in all of the categories. Specifically, we hypothesized the collectivistic and high-contact orientation of the Russian cosmonauts would cause them to be lower in idiocentric behavior, higher in sociocentric behavior, and more easily influenced in terms of behavioral outcomes.

The results indicated partial support for this hypothesis. There were several types of idiocentric behaviors that were significantly different and in the predicted direction when comparing Americans and Russians. For example, Americans were more likely to be dominant, regulate the flow of the conversation, and express strong opinions when interacting with a Russian counterpart. In addition, many other categories of idiocentric behavior also were higher for Americans, albeit not significantly so. However, contrary to expectations, Russians scored significantly higher in one aspect of idiocentric behavior: focusing on the task. Also contrary to expectations, no significant differences were found between Russians and Americans in the various sociocentric categories. Despite this fact, the average frequency counts for the sociocentric indicators were in the hypothesized direction. Finally, the behavioral outcomes were consistent with predictions (i.e., Russians changed their behavior in response to American crew member influence more frequently than the reverse).

Table 1: ICC and Alpha values for the ratings of each of the 20 videos

Video Number	Type of Interaction	ICC
1	Cross-Cultural	.87
2	American-American	.51
3	Cross-Cultural	.48
4	Cross-Cultural	.55
5	Cross-Cultural	.66
6	American-American	.48
7	American-American	.73
8	Cross-Cultural	.70
9	American-American	.67
10	Cross-Cultural	.40
11	Cross-Cultural	.66
12	American-American	.53
13	American-American	.43
14	American-American	.41
15	Cross-Cultural	.49
16	Cross-Cultural	.60
17	American-American	.71
18	Cross-Cultural	.83
19	American-American	.42
20	American-American	.68

Table 2: American-Russian Behavioral Comparisons Across All 20 Videos

	Condition			
	Intracultural		Intercultural	
Individual Behavior	(A) American	(B) American	(A) Russian	(B) American
Idiocentric				
Expressiveness	4.73	5.20	5.93	5.37
Dominance	0.70	1.37	**0.33***	**0.70***
Initiate Action	3.97	4.90	2.93	3.13
Aggressiveness	0.33	0.37	0.03	0.13
Logical / Systematic	0.37	0.70	0.07	0.07
Regulate Flow	1.10	1.10	**1.30****	**2.80****
Focus on Task	0.57	0.70	**0.37***	**0.13***
Strong Opinion	0.97	0.87	**0.30***	**0.73***
Sociocentric				
Agreeable	2.37	2.00	1.00	0.80
Avoid Argument	0.10	0.20	0.10	0.00
Shift Opinion	0.03	0.13	0.00	0.00
Eye Contact	1.73	1.30	2.07	1.60
Outcomes				
Changed Behavior	1.90	1.97	**2.13***	**1.63***
Influenced Other Crewmember	2.17	2.20	1.67	2.03

NOTE: Chi-square for bold cell counts by condition significant at $^{**}p < .05$; $^{*}p < .10$.

Conclusions & Implications

Based on their differing standings in terms of high- and low-context communication (Hall, 1976), we expected American and Russian crew members to demonstrate different ways of communicating. To ensure that we were not simply capturing individual differences in extraversion, we included a control group of 10 videos that featured two Americans conversing with one another. Consistent with our expectations, there were no significant differences in any of the communication behaviors or outcome variables in the American-American videos.

Among the videos featuring Russians and Americans interacting, however, there were some significant differences. These results provide some evidence that there are in fact differences in the way Russian and American crew members communicate. In general, Americans tended to be much more straightforward, direct, opinionated, and dominant when interacting with their Russian counterparts. The Russians, in turn, tended to fall on the low end of these categories and also were higher (although not significantly so) in sociocentric tendencies such as being agreeable and making eye contact.

Should these findings generalize to the astronaut population at large, they will have serious implications for the longer duration space flight expected in the future. On missions where the crew has greater autonomy, crew members with more direct communication styles (i.e., Americans) may exert a greater influence on the order of tasks or manner in which work is performed. Alternatively, those with greater sociocentric tendencies (i.e., Russians), may forge valuable relationships with others and foster heightened levels of cooperation given their more relationship-oriented communication style. Simply the fact that there are differences between the two may hamper efficiency, at least in the beginning as the two parties get accustomed to the alternative ways of communicating.

One unexpected finding was that the Russians were actually significantly higher on being concerned with finishing the task (an idiocentric indicator). In our post-hoc thinking, we reasoned that this may not be so surprising when one considers the joint effects of the post-Soviet Union culture shift as well as current compensation practices of the Russian space agency. In the Soviet Union, there were no differences in terms of pay or recognition. In Russia today, however, there is a sharp contrast and people are able to recognize that there is a clear and transparent link between working hard and gaining intrinsic and extrinsic rewards. This is further underscored by the Russian Space Agency's policy of paying cosmonauts for quality of performance and the number of goals that were met. Astronauts, in contrast, are paid a set amount regardless of mission outcomes. This policy is likely to underlie the significant and unexpected finding that Russians are more focused on the task than Americans.

We must also address the fact that several of the categories failed to show differences between Russians and Americans. One potential explanation for the null findings may be that such a diverse and isolated team such as the mission crews on ISS may find themselves forming an emergent team culture rather than conforming or clashing based on country culture. Because multicultural team members have so little in common from a national culture perspective, new norms, rules, and expectations are sometimes created for a specific team that do not align with either Russian or American cultures per say (Earley & Mosakowski, 2000). In this way, the creation of a strong emergent culture may cause new norms to be adopted, increased trust, and more efficient performance. Alternatively, it also may be that these astronauts are not representative of their average cultural norms. Just as astronauts in general are exceptional in terms of ability and motivation, they may not be representative in terms of natural preferences and may be more culturally savvy and open to experience than their average national counterpart. In general, however, we believe these results highlight that at least modest differences in communication style do occur. Furthermore, because astronauts have the ability to turn the on-board cameras on and off, cross-cultural issues negatively impacting communication may actually be more prevalent than this sample of videos would suggest.

Study 3

Objective

The objective of this last study was to see whether recent work in politeness analysis (e.g., Miller, Wu & Funk, 2009; Miller, Galunder & Ott, 2010, building on Brown and Levinson, 1987) could predict and explain cultural communication behaviors at a finer level of granularity than prior studies. Similar to the above studies, we again utilized video footage that included interactions between astronauts and cosmonauts.

Method

To examine how Brown and Levinson's theory of politeness could inform cross-cultural interactions of astronauts aboard the ISS, a cross-cultural video was selected that included both American and Russian counterparts. This video was selected based upon its content of cross-cultural interactions and its richness in exemplifying how interactions may be explained and even quantified by this theory of politeness. As with the other studies in this report, this video was pulled from the VAMS online film repository.

"Politeness" is frequently regarded as having little applicability in the work place, but in socio-linguistics, politeness refers to the range of indirect communicative behaviors used to convey, signal, and manipulate a wide range of social parameters. Brown and Levinson's (1987) politeness theory claims that each encounter between individuals has the potential to "threaten the face" of those involved—and that the goal of most participants is generally to maintain the social status quo within the interaction. Their theory provides an understanding of interactions between people based on different levels of imposition, familiarity, and power/authority, as well as other differences including status, gender, age, and socio-economic status. When interacting with someone for the first time, certain cultural cues provide information to both parties as to which communication style and level of politeness is required. These cues are very important and many represent ideologies that are engrained within a culture. However, when considering interactions across different cultures, Brown & Levinson's theory of politeness provides insight as to how communications between individuals that understand little of each other's culture may experience difficulty during an interaction. In this instance, the rules of politeness are still relevant (i.e., preserving face), however, the appropriate communication styles, mannerisms, and behaviors that are required are less clear as cultural variations may be great—and misinterpretations are entirely possible. Even long-established relationships may be subject to subtle but persistent misunderstandings about the intended social cues which each side conveys. Thus, Brown and Levinson's theory provides a model to understand how cross-cultural interactions may lead to greater imbalance as perceived by each party that is participating in the interaction.

To review, Brown and Levinson's politeness model comprises two main components: face threat and redressive strategies (Brown & Levinson, 1987). All social actors have "face," which is "the positive social value a person effectively claims for himself" (Goffman, 1967, p.5). Face can be "saved" or lost, and it can be threatened or conserved in interactions. Face threat results from any behavior and/or act that thwart one's will or sense of self-determination or self-worth or, loosely, that threaten one's ego. Face represents the public self image that one portrays to the external world and what every adult tries to project (Brown & Levinson, 1987). Thus, a face-threatening act damages the face of the addressee by acting in opposition to the wants and desires of the other. The degree of face threat in an interaction, in Brown and Levinson's model, is a function of power difference (level of authority), social distance (level of familiarity), and imposition (level of interference the requester's goal interferes with the hearer's goal).

Redressive strategies are what we normally think of as explicit "politeness" behaviors—techniques such as "please" and "thank you" used to reduce the negative face threat that is perceived by an individual. There are two type of redressive strategies: positive, in which one seeks to minimize the threat (e.g. offer or promise = "I'll wash the dishes, if you'll vacuum the floor" or include both speaker and hearer = "If *we* help each other…"), and negative, in which one emphasizes avoidance of imposition on the hearer (e.g., minimize the imposition, "If it's not too much out of your way…" or apologize, "I'm sorry; I know it's a lot to ask…"). In Brown and Levinson, every interaction has a face threat potential, but these are perhaps most salient in requests or directives. Miller and colleagues (e.g., Miller, et al., 2009) have advocated computing the net "imbalance" in interactions (approximately, the perceived politeness) by comparing the amount or value of the face threat present to the value of the redress used. The equation is as follows:

Imbalance = Redressive Strategy – Face Threat*

*where Face Threat = Power Difference + Social Distance + Imposition

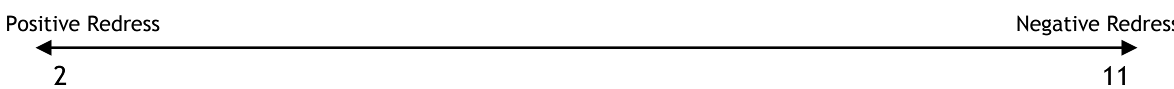
Thus, for an interaction, the type and value of redressive strategies used may balance the face threat that is perceived by both parties, resulting in perception of a "nominal" or just normally polite interaction. Alternatively, if substantially more redress is used than there was face threat, the interactions will be perceived as overly polite, while if substantially less redress is used, it will be perceived as rude. Again, each individual's perception of the amount of face threat present and redress used may differ (especially likely in cross cultural interactions), resulting in differing perceptions of whether the interaction was polite, rude, or nominal.

Finally, Brown and Levinson imply that, in interaction, people may have a wide range of motives for employing politeness or rudeness (meaning, more or less redressive value than they perceive as necessary). In general, though, with all other things being equal, we strive to maintain the social status quo—which means using an amount of redress that in context (that is, given the power, familiarity/social distance and imposition relationships present) will be perceived as nominal. In fact, whenever deviations from nominal interactions are perceived, the observers can be expected to seek explanations in the context, which will explain away the deviation by making it balance. For example, rudeness may be perceived as urgency (that is, justified by reduced imposition) or by increased familiarity (reduced social distance), etc.—each of which would necessitate less redressive behavior than the observer may have otherwise expected.

In this specific case study, a quantitative version of the above model was applied to the selected video interaction. Each stage of the communication was charted from both the Russian and American counterpart's perspectives. Thus, we can see how the imbalance equation, as demonstrative of Brown and Levinson's theory of politeness, accounts for differences in perceptions of an interaction between two individuals and may explain their subsequent behaviors over time (Miller et al., 2009).

In terms of scaling for each of the components of face threat and redressive strategies, the following ranges were used:

- Redressive Strategy (negative and positive redress are viewed on a continuum):
    - Positive Redress: (lowest category, received 2 points)
    - Negative Redress: (highest category received 11 points; please refer to Brown and Levinson, 1987, for a complete listing of the redress categories).



- Face Threat
    - Power Difference
        - Commander talking to subordinate = 1
        - Two crew members talking to each other = 5
        - Subordinate talking to Commander = 11
    - Social Distance
        - Mission Duration/10 = (where clip falls will be representative of phase of mission- most direct indicator of familiarity)
    - Imposition
        - Working together toward same goal = 1

- Request to meet personal goal that doesn’t interfere with receiver’s goal = 5
- Request to meet personal goal that interferes with receiver’s goals = 11

Results

To provide some background information about this specific clip, the American astronaut is exercising on a bike aboard the ISS. He is about 20 minutes into his exercise (we assume that the standard exercise is one hour at a time) when the Russian cosmonaut (and it is worth noting, he is also the commander of this Expedition) comes to him to make a request about labeling some food containers (a request, as it turns out, that originated from mission control).

Utterance 1:

Russian Cosmonaut (B), “[Astronaut’s Name]?”

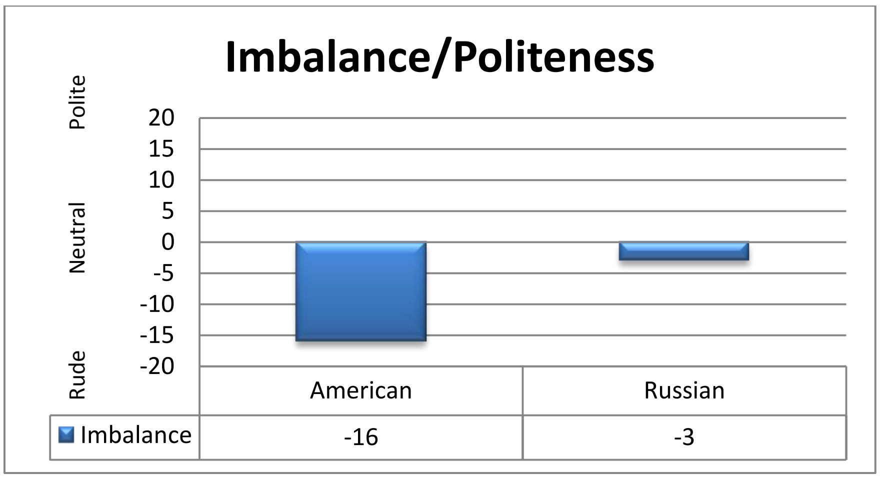
American Astronaut (A), “Yes, [Cosmonaut’s Name]”

B, “Can you check… uh… the numbers on the containers… food containers?”

From the Russian’s perspective, we assume he believes that he (B) is making a pretty standard request, not a large imposition; B used an in-group identity marker by calling A by name, and asked A an indirect question (noted with the hesitancy… “uh…”). However, from A’s perspective, the imposition was substantially larger (as expressed later in A’s explanation, this is a task he has been fighting to avoid doing, as well as coming at a time where it interrupts his cycling), and the redressive strategies that were used by the cosmonaut were either not perceived as such (the hesitant “…uh”) and/or were inadequate to compensate for the face threat resulting from the increased imposition stemming from interfering with A’s personal goal of completing his exercises on the bike.

Table 3: Utterance 1

	Redressive Strategy	Power Difference	Social Distance	Imposition	Face Threat	Redressive – Face Threat	Imbalance/ Politeness
American	0	1	4	11	16	**-16**	**-16**
Russian	7	1	4	5	10	**-3**	**-3**



Graph1: Imbalance 1

Utterance 2a:

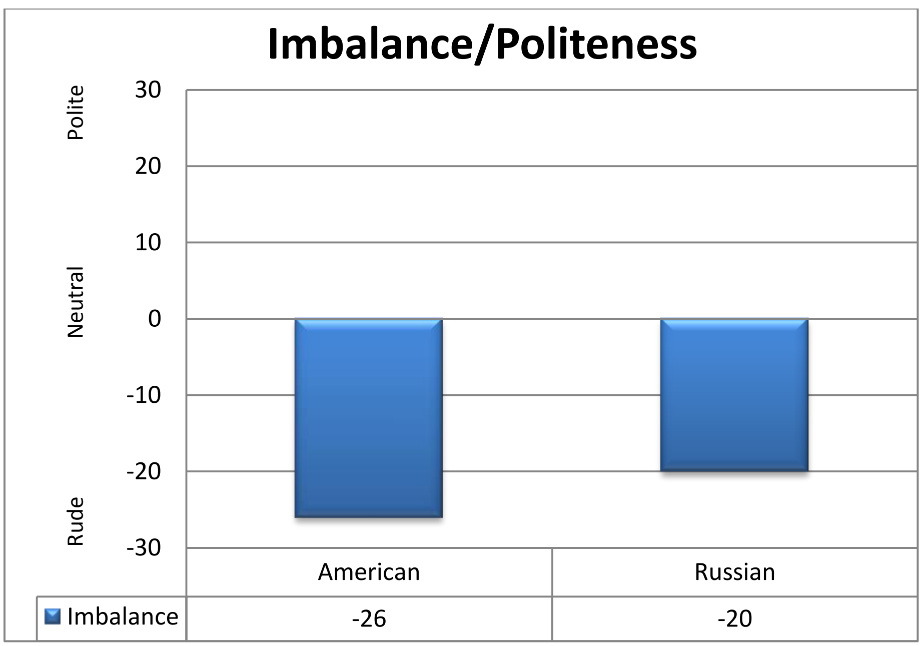
Astronaut: "No… I'm not going to do that."

[Slight pause]

The American's initial response was perceived rude by both parties; the American was responding to the Russian (as a subordinate to a commander) using no redressive strategy—as bluntly as possible. Coming after a request from a commander, this blunt refusal without explanation might even be taken as having "negative redress"—that is, adding to the face threat of the fact of the rejection—though we haven't scored it that way. Note also the lack of any in-group identity strategy (A says "I'm not going to do that" not "We don't have to do that"). In this instance, the power difference, social distance, and imposition of the initial request have not changed—though note that since the subordinate is now addressing his commander, the power difference value changes. We will have more to say about this below, but note too the differences in perceptions of the two individuals. Since A perceived B as being rude to him in the initial interaction, his rudeness here is at least somewhat justified as a response, but since B perceived (and presumably intended) no rudeness in his initial interaction, A's rudeness here is unexpected and extreme—perhaps signaled by the pause and B's "being at a loss for words" to respond.

Table 4: Utterance 2a

	Redressive Strategy	Power Difference	Social Distance	Imposition	Face Threat	Redressive – Face Threat	Imbalance
American	0	11	4	11	26	**-26**	**-26**
Russian	0	11	4	5	20	**-20**	**-20**



Graph2a: Imbalance

Utterance 2b:

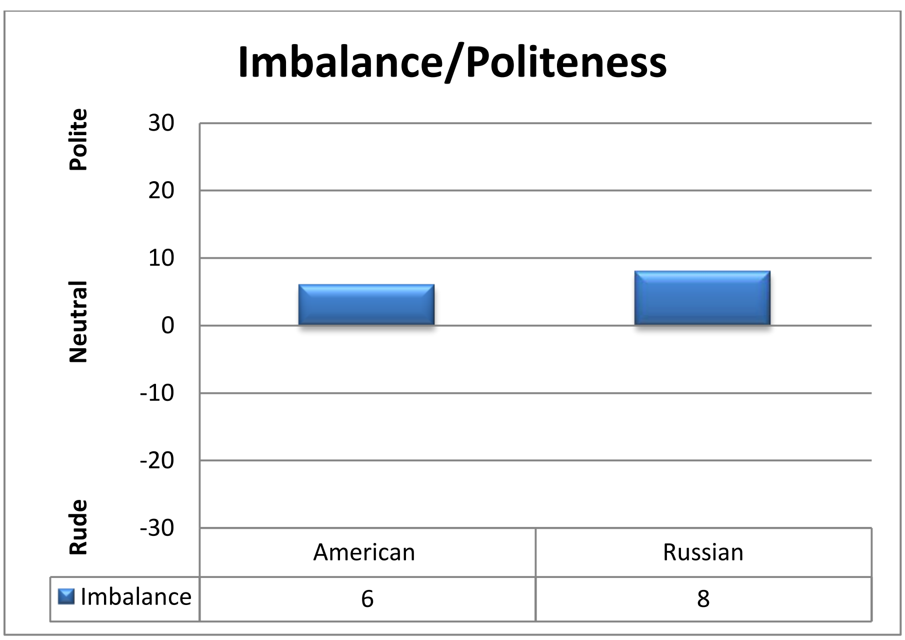
[After slight pause]

A: "My commander said we weren't going to waste our time on that."

Perhaps realizing that he has overstepped the polite boundaries of their relationship, A takes a conversational step out of turn and begins a redressive process by offering an explanation. Note too the use of the in-group identity marker "we." Finally, the reference to 'his commander' is odd. It is an example of the negative redressive strategy to impersonalize or shift the blame for an action onto a third party or general rule—though that may be a problematic strategy here. By "my commander" A is in fact referring to B, hence we've given it less than full value (it is also possible that A is referring to a previous mission commander). Note also that the topic of conversation remains the initial request to check the food containers, so we have retained the initial imposition value.

Table 5: Utterance 2b

	Redressive Strategy	Power Difference	Social Distance	Imposition	Face Threat	Redressive – Face Threat	Imbalance
American	10	11	4	11	26	**-16**	**-16**
Russian	10	11	4	5	20	**-10**	**-10**



Graph2b: Imbalance

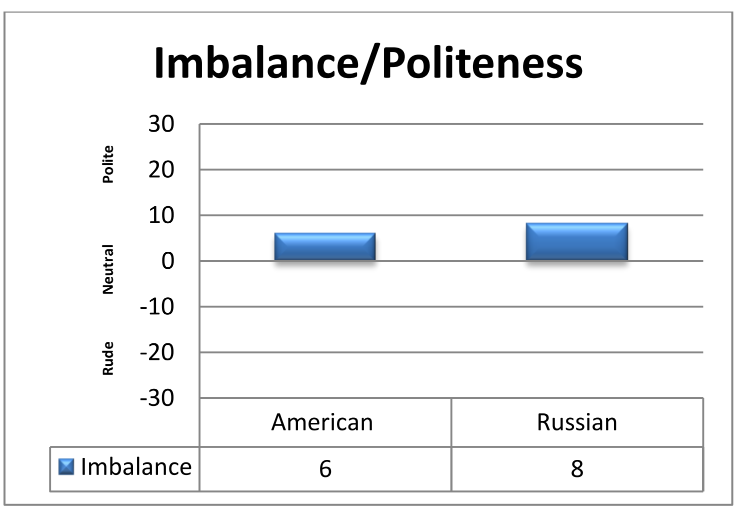
Utterance 2c:

R: "Yeah, yeah, yeah…. They have some disagreement, the Russian and Americans."

B agrees with A here and picks up on the theme of some external source for the order (probably ground control)—thereby also impersonalizing the imposition. Also, the reference to "the Russian and Americans" as a group other than "us" can be seen as an in-group identity assertion. Finally, the topic of conversation is beginning to shift here from the initial request to check the food containers to a discussion of the disagreement. There is no inherent request here and they both share the goal to mitigate this disagreement, so we have dropped the imposition value for this utterance accordingly.

Table 6: Utterance 2c

	Redressive Strategy	Power Difference	Social Distance	Imposition	Face Threat	Redressive – Face Threat	Imbalance
American	13	1	4	1	6	**7**	**7**
Russian	13	1	4	1	6	**7**	**7**



Graph2c: Imbalance

Utterance 3:

Astronaut B: "[The mission control] does not understand… [unclear]

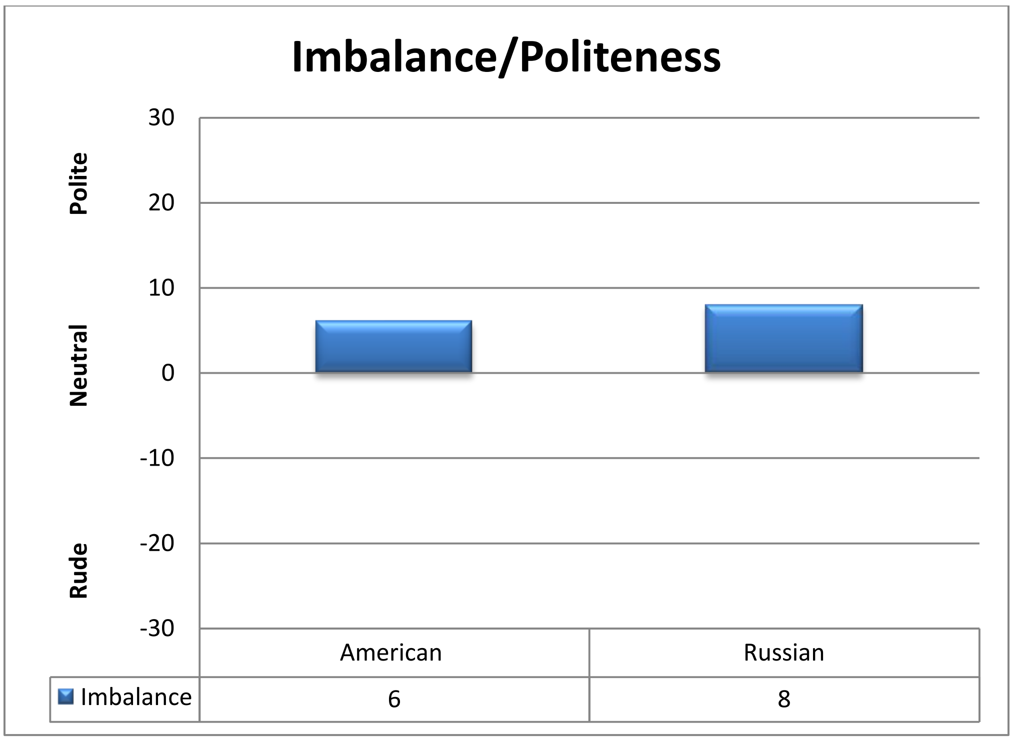
Astronaut A: Switches to Russian to further explain his side of the story (translation included below).

Astronaut A	They know. I counted and told them. I wrote a letter stating exactly what we have. 144. (Unclear) When I counted all of them, I said how many containers we have that are not opened and how many we have that we had opened, and I estimated how much is in each bag. And I said 74 or 72 are shared bags.. eh.. containers.
Astronaut B	We have 72 containers total.
Astronaut A	Yes. If we count what we have now in Node
Astronaut B	Yes yes yes. Yes
Astronaut A	But each … They said they want to know the exact serial number of each bag. Remember we spoke about it.
Astronaut B	Yes yes
Astronaut A	I said, it's not necessary. I said we don't normally do that because food is food. Doesn't matter if it's box 1 or 10. I said: "I won't do that. If you want to speak to the Commander… You can speak to the Commander."
Astronaut C	Unclear. We'll have to speak to them again
Astronaut B	They probably have a problem with whether it's one or the other. That's why they want…
Astronaut B	Yes. It's not our problem. It's the problem they have on the ground
Astronaut A	Unclear (sounds like: "Their problem is to figure out")… how much we need and how to deliver it up here. After that - no.
Astronaut B	But you know there are VERY many people who are very involved with us here

At this point in the interaction, it is very clear that the American perceives his own rudeness in talking to the cosmonaut; thus, he goes into a two-minute plus explanation about why he is not willing to label the food containers. In addition, he also switches to speaking to Russian, a sign of deference (these actions increase the redressive strategy score in this extended utterance, which we have collapsed into a single value for ease). Furthermore, since this interaction is all about the degree of imposition associated with the requested action, along with whose imposition it should be and how important it is, it seems reasonable that their difference in perception of the request should be mitigated—and we have reflected that fact in the imposition value (while preserving some difference to reflect what we perceive to be a continued preference on B's part for satisfying ground control's request vs. a preference for avoiding it on A's part.)

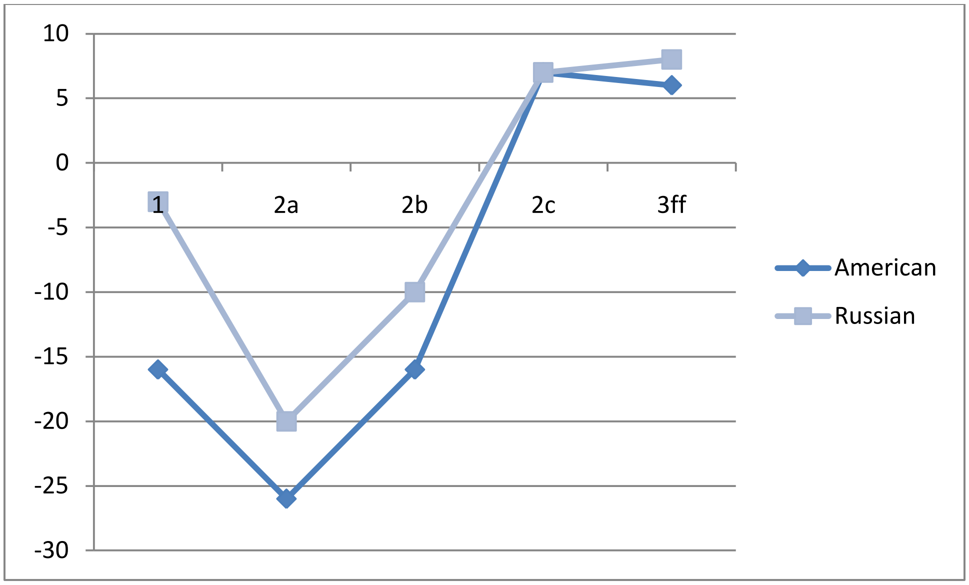
Table 7: Utterance 3

	Redressive Strategy	Power Difference	Social Distance	Imposition	Face Threat	Redressive – Face Threat	Imbalance
American	30	11	4	9	24	**7**	**6**
Russian	30	11	4	7	22	**7**	**8**



Graph3: Imbalance

Thus, we see how Imbalance/politeness can change over time and may differ among the parties involved, especially when those individuals are from different cultures. The graph below depicts how imbalance did change over time from each individual perspective:



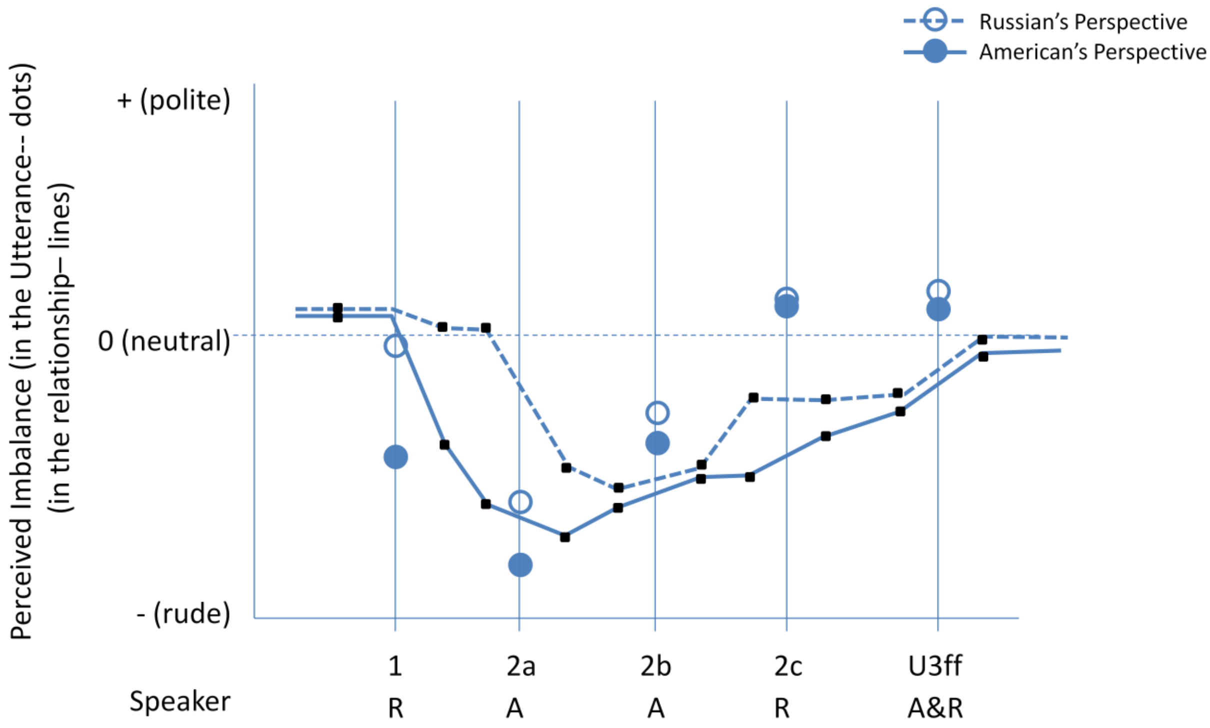
Graph4: Imbalance over Time

Note that this graph depicts only a quantification of the imbalance in politeness that each participant perceived in their conversation over time. Numbers below zero convey relative perceived rudeness, while those above zero convey increased politeness. While this graph is not meant to convey their overall attitude toward each other, much less a notion of team cohesion on an utterance by utterance basis, it is likely that sustained interactions of this sort will impact those parameters. Instead, we see here that R began the conversation intending (and perceiving) only a mild imbalance—consistent with a routine request, adequately redressed, especially given his role as commander. By contrast, A perceived substantial rudeness—and then responded with a statement that was even ruder. This unexpected (from R's perspective) rudeness at utterance 2a, was extremely sharp. Thereafter, we see consistent attempts to repair the situation by increased levels of politeness and a narrowing of the gap in perceptions between A and R.

Maintaining the Relationship

While the graph above depicts perceived imbalance over time is not meant to depict the impact of the politeness actions on the relationship between the two participants, in principle, both participants are likely concerned about the relationship and doing things to shape and repair it. In keeping with Brown and Levinson, the graph depicts the perceived imbalance at each utterance by each party, probably more important for long-term relationships is the "residual imbalance" or cumulative effect of multiple perceived polite and impolite interactions. This is, quite possibly, why A changes course at utterance 2b and into utterance 3ff from increasing the level of rudeness to one of offering increased (even, in the moment in 3ff, over-) politeness. But this is a step beyond what Brown and Levinson define and would necessitate substantial future work to model and quantify.

Nevertheless, we can convey our intuitions about what is going on in the interaction by overlaying the impact of the various utterances on the individuals' perceptions of their "relationship." We might approach greater concreteness (and a more nearly testable model) if we were to substitute more concrete attributes such as affect, trust, perceived cohesion, etc. for "relationship," but it is likely that the graphs would be similar.



Graph5: Perceived Imbalance

This figure depicts the hypothesized effect that the utterances have on "their relationship" as each of them perceives it. To depict this, we have constructed a different graph than the one presented above. Here, perceived imbalance (taken from the tables above) is graphed for each utterance via the hollow and filled circles for A and B respectively. Connecting these points directly would produce the same figure as graphed above—their perceived imbalance of each utterance. The solid and dashed lines, however, depict the perception that we believe each individual would have about the effects of the utterance on their relationship. We have placed an inflection point (in the form of a small black square) both before and after each utterance point to reflect the intent (or expected intent) of the utterance (before) and the perceived effect of it (after). Also, just like the perceived imbalance of the utterances themselves, perceptions of the relationship can differ between people. Also, while utterances influence (or "pull on") the relationship (the lines), relationships have a momentum of their own and it is difficult for a single utterance to completely change that momentum. Thus, while the force of an utterance pulls a relationship in its direction, it must contend with the prior momentum of the relationship.

To illustrate, let's consider the conversation in conjunction with the information provided above. Prior to the first utterance in this sequence, it can be assumed that A and B have a reasonably good relationship and that the expectation of any new communication would be to retain and further that relationship. Hence, both parties' expectations of the first utterance were that it remains in keeping with their neutral-to-good feelings toward each other. Instead, after U1, A has perceived a very rude comment (an inconvenient request without any significant redress) and his sense of their relationship is greatly decreased. By contrast, B thinks that he, at most, made a minor imposition on his friend A and the relationship should not be greatly affected.

Before utterance 2a, therefore, A intends a response on par with the rudeness he just perceived, perhaps conveying his affront at the rudeness he received and, therefore, the fact that this request from B was much more imposing than B intended it to be. On the other hand, B also expects an utterance in keeping with his perception of no significant change to their positive relationship.

Instead, A makes the utterance 2a and it is much ruder than B was expecting—hence the precipitate drop in his perception of the relationship. Astronaut A perceives maybe a bit more of a drop than he intended—perhaps due to B's pause and "loss for words." For that reason, he then decides that things ought not to stay at this low level and takes a step "out of turn" to begin attempts to repair them. Before utterance 2b, he intends to improve the situation, while B may expect continued declines.

Astronaut A makes utterance 2b, with an increase in politeness, and perceives that it has helped things somewhat. Astronaut B, too, sees this as an attempt to repair the situation and his perception of the relationship improves as well—though both are aware that it has taken a blow. Astronaut B then intends utterance 2c as a further attempt to repair, expecting it improve things further. Astronaut A, though, is unsure how his utterance 2b has gone over and unsure what B's next response will be and what impact it will have on the relationship.

With B's comparatively polite utterance 2c, though, Astronaut A is reassured and the perceptions of both A and B is that the relationship is being repaired. Both parties then enter into continued repair actions (with A bearing the brunt of the work) in utterances 3ff, as their perception of their relationship slowly returns back toward where it was before (perhaps a bit suppressed) and back toward synchronization with each other.

Though hypothetical, this exercise represents the kinds of implications that might eventually be drawn from politeness variations to their implication on aspects of team relationships and cohesion. These differences in perception arise both from individual differences (in both personality and current attitude and knowledge) and from cultural differences. The ability to track such differences in interpretation would, of course, be very valuable in managing, diffusing, and even perhaps predicting mismatches during long-duration missions.

Conclusions & Implications

Although preliminary and of limited scope, this case study exemplifies how cross-cultural differences may be influenced by expectations of politeness and, in turn, may influence the outcome of an interaction. In addition, this case study also supports the use and application of Brown and Levinson's model of politeness when examining interactions among individuals. The importance of understanding how cross-cultural differences may influence interactions is especially important when considering long-duration missions. Indeed, effective communication is a major component of teamwork and will be vital to the success and completion of a long-duration mission. Thus, it is essential that countermeasures (including selection and training) incorporate, as requirements of training opportunities, acquiring a thorough understanding and knowledge base about each other for optimal teamwork to be achieved. It is important to note however, that this application spans more than mere ethnic cultures. Cultures can encompass many different levels including job role, education, or organization and need to be included when considering effective teamwork and team performance as well.

As with any study, this case study has many limitations that need to be discussed. First, while we applied a quantification metric, this was in no way a systematic examination of cross-cultural interactions and thus does not provide quantitative, statistical evidence of the findings that were described. Secondly, the video was selected with the intention of exemplifying a difficult interaction between crew members; this in no way is representative of a crew's entire experience and/or all of their interactions aboard the ISS. However, if warranted, a more thorough examination of more cross-cultural interactions could provide insight into the frequency in which these types of interactions do occur. Thirdly, these results may not be generalizable to the astronaut population as a whole. Brown and Levinson's model was utilized to exemplify that the components of the framework could be applicable to interactions aboard the ISS, but do not relate to a pattern of behaviors, communication styles, and reactions to the astronaut population on a larger scale.

While maintaining the limitations outlined above, the evidence presented by this case study still provides support for the importance of cross-cultural interactions. To ensure effective teamwork and team performance among crew members for a long period of time (as will be necessary for a long-duration mission), it is essential that the crew members can work effectively with each other and can communicate, behave, and react in ways that support the team dynamic. As one possibility, future investigations should consider a quantitative approach of applying Brown and Levinson's (1987) model to the examination of cross-cultural interactions to determine how selection and training processes can be improved to ensure optimal team performance and effective teamwork.

General Discussion

These three studies collectively suggest that differing cultural backgrounds do, at a minimum, impact the way that crew members communicate with one another and also result in different verbal and nonverbal behaviors. In the first study, we saw several examples of instances where factors such as power distance, individualism and collectivism, and culture-specific knowledge impact how people interact with one another aboard the space station and also impact whether one's true meaning is conveyed. In the second study, we saw that empirical differences emerged in communication style, such that Russians were more indirect and agreeable (i.e., less idiocentric), whereas Americans were viewed as somewhat more domineering and opinionated. Finally, in the third study, we examined how differing cultural preferences regarding positive and negative politeness can impact communication.

There are several ways that selection and training can be augmented to create better functioning cross-cultural long-duration crews in light of these findings. For example, Davison (1994) discussed the creation of a high-performing multicultural team beginning at the selection stage. She stated that such teams can be created by (a) choosing the right mix of people (with regards to cultural background as well as personality, skills, and knowledge); (b) removing constraints of strict rules or bureaucracy; (c) sharing the goals and objectives with the team and involving them in the management process; (d) appreciating the influence of nationality and the unique perspectives it can bring (offer training to augment or highlight this if necessary); and (e) choosing the right leader and making all members accountable for outcomes (and assuring that no one is excluded; Davison, 1994).

Another approach organizations can take is to increase the cross-cultural communication competency of team members to ensure better verbal and nonverbal cross-cultural communication (Spitzberg, 1983). Previous research suggests that this competency requires sufficient knowledge of other cultures as well as ample inquisitiveness and patience with multicultural team members (Kealey & Protheroe, 1996; Spitzberg, 1991). Others also emphasize the use of skills such as empathy, charisma, and the ability to manage anxiety in the face of ambiguity (e.g., Gudykunst, 1998). It is important to foster this aptitude as it can have a direct impact on multicultural team performance. Specifically, team members high in cross-cultural communication competency can express themselves more clearly and efficiently with members of other cultures when engaged in work tasks (Matveev & Nelson, 2004).

The Cross Cultural Communication Competency model builds on previous conceptual work (Abe & Wiseman, 1983; Cui & Awa, 1992) to provide a theoretical framework of this competency. This model distinguishes between four facets: interpersonal skills, team effectiveness, cultural uncertainty, and cultural empathy (Matveev, 2002; Matveev & Nelson, 2004). The interpersonal skills dimension reflects both an understanding and comfort when interacting with people who have different communication styles as well as when resolving disputes. The team effectiveness facet reflects the ability to clearly communicate team goals and roles to other team members. Cultural uncertainty is primarily demonstrating patience with other team members and with unfamiliar customs. Finally, cultural empathy involves seeing things from the perspective of another culture and behaving accordingly. This final dimension also reflects attitudes about working on a multicultural team, such as refraining from value judgments like good or bad and right or wrong (Matveev & Nelson, 2004). It is important to facilitate the development of each of these facets in crew members to ensure effective communication and conflict resolution on multicultural long-duration missions.

In conclusion, we believe that these three studies jointly suggest that cross-cultural communication issues, in fact, do occur on board the ISS and have the ability to impact team work and task effectiveness. Because all three studies used the same data source, replication and extension of these findings is warranted. Future work should include interviews with past, current, and future astronauts from a variety of cultures to ascertain their perspective on how these issues impact crew effectiveness and how they might best be mitigated. Further studies could be conducted by using content analysis techniques on astronaut biographies, journals, and other historical documents. Finally, an actual study aboard the ISS that records conversations on a continuous basis would be most helpful for ensuring that valuable data points are not being edited out. Future work also should explore the feasibility and effectiveness of implementing cross-cultural competency training as part of the ASCAN or mission training flow.

References

Abe, H., & Wiseman, R. (1983). A crosscultural confirmation of the dimensions of intercultural effectiveness, *International Journal of Intercultural Relations, 7*, 53–67.

Bantz, C. R. (1993). Cultural diversity and group cross-cultural team research. *Journal of Applied Communication Research*, *February*, 1–20.

Berdiaev, N. (1990). *Knowing yourself*. Moscow: DEM.

Betancourt, H., & Wiener, B. (1982). Attributions for achievement-related events, expectancy, and sentiments: A study of success and failure in Chile and the United States. *Journal of Cross-Cultural Psychology, 13*, 362–374.

Bollinger, D. (1994). The four cornerstones and three pillars in the "House of Russia" management system. *Journal of Management Development*, *13*, 49-54.

Boyd, J. (2005). Cultural factors and the International Space Station. Aviation*, Space, and Environmental Medicine, 76*, B135-B144.

Boyd, J. E., Kanas, N., Gushin, V. I., & Saylor, S. (2007). Cultural differences in patterns of mood states on board the International Space Station. *Acta Astronautica, 61*, 668-671.

Brett, J. M., Adair,W., Lempereur, A., Okumura, T., Shikhirev, P., Tinsley, C., & Lytle, A. (1998). Culture and joint gains in negotiation. *Negotiation Journal, 14*, 61-86.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Camiah, N., & Hollinshead, G. (2003). Assessing the potential for effective cross-cultural working between 'new' Russian managers and western expatriates. *Journal of World Business*, *38*, 245-261.

Cui, G., & Awa N. E. (1992). Measuring intercultural effectiveness: An integrative approach. *International Journal of Intercultural Relations, 16*, 311–28.

Davison, C. S. (1994). Creating a high performance international team. *Journal of Management Development*, *13*, 81-90.

Earley, P. C., & Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal, 43*, 26-49.

Fiske, A., Kitayama, S., Markus, H., & Nisbett, R. (1998). The cultural matrix of social psychology. *The handbook of social psychology, Vols. 1 and 2 (4th ed.)* (pp. 915-981). New York, NY US: McGraw-Hill.

Goffman, E. (1967). *Interactional Ritual*. Chicago: Aldine.

Gudykunst, W. B. (1998). Applying Anxiety/Uncertainty Management (AUM) theory to intercultural adjustment training. *International Journal of Intercultural Relations, 22*, 227–50.

Hall, E. & M. Hall. (1990). *Understanding cultural differences.* Yarmouth, Maine: Intercultural Press.

Hofstede, G. (1980) *Culture consequences: International differences in work-related values.* Beverly Hills, CA: Sage.

Hofstede, G. (2001). Culture's consequences: Comparing values,behaviors, institutions and organizations across nations (2nd ed.). Thousand Oaks, CA: Sage.

Hofstede, G., & McCrae, R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research: The Journal of Comparative Social Science*, *38*, 52-88.

Jandt, F. E. (2006). Dimensions of culture. In F. E. Jandt, *An introduction to intercultural communication: Identities in a global community* (5th ed., pp. 159-181). Thousand Oaks, CA: Sage Publications.

Kanas, N., Sandal, G., Boyd, J., Gushin, V., Manzey, D., North, R., Leon, G., Suedfeld, P., Bishop, S., Fiedler, E., Inoue, N., Johannes, B., Kealey, D., Kraft, N., Matsuzaki, I., Musson, D., Palinkas, L., Salnitskiy, V., Sipes, W., Stuster, J. & Wang, J. (2009). Psychology and culture during long-duration space missions. *Acta Astronautica, 64*, 659-677.

Kanter, M. R., & Corn, R. (1994). Do cultural differences make a business difference?. *Journal of Management Development*, *13*, 5-23.

Kealey, D. J., & Protheroe, D. R. (1996). The effectiveness of cross-cultural training for expatriates: An assessment of the literature on the issue. *International Journal of Intercultural Relations, 20*, 141–65.

Kirby, L., & Barger, N. (2009). Negotiating group dynamics. *Diversity Executive*. Retrieved from http://www.diversity-executive.com/article.php?article=473

Kowner, R., & Wiseman, R. (2003). Culture and status-related behavior: Japanese and American perceptions of interaction in asymmetric dyads. *Cross-Cultural Research: The Journal of Comparative Social Science, 37*, 178-210.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Lewis, R.D. (1996). *When cultures collide: Managing successfully across cultures*. London: Nicholas Brearley Publishing.

Matveev, A. V. (2002). *The perception of intercultural communication competence by American and Russian managers with experience on multicultural teams* (Doctoral dissertation). Ohio University, Ann Arbor, MI.

Matveev, A., & Nelson, P. (2004). Cross cultural communication competence and multicultural team performance: Perceptions of American and Russian managers. *International Journal of Cross Cultural Management*, *4*, 253-270.

Maznevski, M. L. (1994). Understanding our differences: Performance in decision-making groups with diverse members. *Human Relations, 47*, 531–52.

McSweeney, B. (2002) Hofstede's model of national cultural differences and their consequences: A triumph of faith– a failure of analysis. *Human Relations*, 55(1), 89-118

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology, 46*, 961–978.

Miller, C., Wu, P. and Funk, H. (2008). A Computational Approach to Etiquette: Operationalizing Brown and Levinson's Politeness Model. IE*EE Intelligent Systems, 23*(4), 28-35.

Miller, C., Galunder, S. and Rye, J. (2010). Politeness in social networks: Using verbal behaviors to assess socially-accorded regard. In *Proceedings of IEEE Social-COM 2010.* Minneapolis, MN; August 20-22.

Molinsky, A. (2005). Language fluency and the evaluation of cultural faux pas: Russians interviewing for jobs in the United States. *Social Psychology Quarterly*, *68*, 103-120.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*, 949–971.

Morrison, T., Conaway, W. A., & Borden, G. A. (1994). *Kiss, bow or shake hands*. Holbrook, Massachusetts: Adams Media Corp.

Naval Historical Center. (2003). Astronauts: Chronology of Space Missions Involving U.S. Navy and Marine Corps Crew Members, 1961- April 1981. Department of the Navy.

Oberg, J. (2007). International Space Station. from http://www.worldbookonline.com/wb/Article?id=ar279523

Pekerti, A. A. (2005). Cross-cultural perceptions in the leadership process: Theoretical perspective on the influence of culture on self-concepts and manager-worker attributions. *Thunderbird International Business Review, 47*, 711-735.

Pekerti, A. A., & Thomas, D. C. (2003). Communication in intercultural interaction: An empirical investigation of idiocentric and sociocentric communication styles. *Journal of Cross-Cultural Psychology, 34*, 139-154.

Rajan, M. N., & Graham, J. L. (1991). Nobody's father was a merchant: Understanding the Soviet negotiation style. *California Management Review, 33*, 40-57.

Sandal, G. M., & Manzey, D. (2009). Cross-cultural issues in space operations: A survey study among ground personnel of the European Space Agency. *Acta Astronautica, 65*, 1520-1529.

Schuster, B., Forsterlung, F., & Weiner, B. (1989). Perceiving the causes and failure: A crosscultural examination of attributional concepts. *Journal of Cross-Cultural Psychology, 20*, 191–213.

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In *Individualism and collectivism,* edited by U. Kim, H.C. Triandis, and G. Yoon. London: Sage.

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

Shenkar, O., & Zeira, Y. (1992). Role conflict and role ambiguity of CEOs in international joint ventures. *Journal of International Business Studies, 23*, 55–75.

Smith, G. (2000). The Mercury, Gemini and Apollo Missions from http://apollo-society.org/merc_gem_apollo.html

Snow, C. C., Snell, S. A., Davison, S. C. & Hambrick, D. C. (1996). Use transnational teams to globalize your company. *Organizational Dynamics, 32*, 20–32.

Spitzberg, B. H. (1983). Communication competence as knowledge, skill and impression. *Communication Education, 32*, 323–9.

Spitzberg, B. H. (1991). Intercultural communication competence. In L. A. Samovar and R.E. Porter (Eds), *Intercultural Communication: A Reader*, 6th edn, pp. 353–65.Belmont, CA: Wadsworth.

Triandis, H. C., McCusker, C., Betancourt, H., Iwao, S., Leung, K., Salazar, M. J., et al. (1993). An etic-emic analysis of individualism and collectivism. *Journal of Cross-Cultural Psychology*, *24*, 366-383.

Vodosek, M. (2009). The relationship between relational models and individualism and collectivism: Evidence from culturally diverse work groups. *International Journal of Psychology*, *44*, 120-128.

Watson, O. M. (1970). *Proxemic behavior: A cross-cultural study*. The Hague, Nederlands: Mouton.

Watson, W. E., Kumar, K., & Michaelson, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal, 36*, 590–602.

Watson, A., Monroe, E., & Atterstrom, H. (1984). American and Swedish children's apprehension about communication: A comparative study. *Perceptual and Motor Skills*, *59*, 917-918.

Appendix A

Communication Style Coding Scheme

Within the brief interaction, how many times was/did each of the two individuals:
1. more expressive during the interaction
2. more dominant during the interaction
3. initiate more action during the interaction
4. more aggressive during the interaction
5. more logical and systematic in his or her arguments
6. regulate the flow of the interaction
7. more concerned with finishing the task during the interaction
8. present stronger opinions during the interaction
9. more agreeable to the partner's suggestions
10. avoid arguments on specific issue(s) during the interaction
11. more readily shift opinions during the interactions
12. make or attempt to make more eye contact with his or her partner during the interaction

Additionally, to what extent (1 being not at all, 3 being somewhat, and 5 being extremely)
13. did each individual influence the other person
14. did each individual change his or her behavior during the interaction

Open-Ended Questions:
1. What verbal markers did you look for when coding for the various behaviors? Please give specific examples.

2. What visual or body-language markers did you look for when coding for the various behaviors? Please give specific examples.

Unobtrusive Monitoring of Spaceflight Team Functioning

Literature Review and Operational Assessment for
NASA Behavioral Health and Performance Element

Veronica Maidel, M.S.

Jeffrey M. Stanton, Ph.D.

School of Information Studies

Syracuse University

Final Report Submission, 8/31/2010

For further information about this report, please contact Dr. Jeffrey M. Stanton, jmstanto@syr.edu, 315-443-2879.

1. Table of Contents

Deleted: 60

2. Executive Summary

This document contains a literature review suggesting that research on industrial performance monitoring has limited value in assessing, understanding, and predicting team functioning in the context of space flight missions. The review indicates that a more relevant area of research explores the effectiveness of teams and how team effectiveness may be predicted through the elicitation of individual and team mental models. Note that the "mental models" referred to in this literature typically reflect a shared operational understanding of a mission setting such as the cockpit controls and navigational indicators on a flight deck. In principle, however, mental models also exist pertaining to the status of interpersonal relations on a team, collective beliefs about leadership, success in coordination, and other aspects of team behavior and cognition.

Pursuing this idea, the second part of this document provides an overview of available off-the-shelf products that might assist in extraction of mental models and elicitation of emotions based on an analysis of communicative texts among mission personnel. The search for text analysis software or tools revealed no available tools to enable extraction of mental models automatically, relying only on collected communication text. Nonetheless, using existing software to analyze how a team is functioning may be relevant for selection or training, when human experts are immediately available to analyze and act on the findings. Alternatively, if output can be sent to the ground periodically and analyzed by experts on the ground, then these software packages might be employed during missions as well. A demonstration of two text analysis software applications is presented.

Another possibility explored in this document is the option of collecting biometric and proxemic measures such as keystroke dynamics and interpersonal distance in order to expose various individual or dyadic states that may be indicators or predictors of certain elements of team functioning. This document summarizes interviews conducted with personnel currently involved in observing or monitoring astronauts or who are in charge of technology that allows communication and monitoring. The objective of these interviews was to elicit their perspectives on monitoring team performance during long-duration missions and the feasibility of potential automatic non-obtrusive monitoring systems.

Finally, in the last section, the report describes several priority areas for research that can help transform team mental models, biometrics, and/or proxemics into workable systems for unobtrusive monitoring of space flight team effectiveness.

Conclusions from this work suggest that unobtrusive monitoring of space flight personnel is likely to be a valuable future tool for assessing team functioning, but that several research gaps must be filled before prototype systems can be developed for this purpose.

3. INTRODUCTION

In the context of space flight, teamwork is an essential ingredient in successful missions. A variety of adverse influences may negatively impact the performance of mission teams both on the ground and in flight. Such influences may include physical stressors on the organism such as diurnal disruption, effects of microgravity, injury, or task overload as well as psychological factors such as social isolation, role overload, or interpersonal conflict among team members. Given the importance of team effectiveness, NASA's Behavioral Health and Performance Element (BHP) has identified a need to monitor the functioning of teams, primarily using unobtrusive means. The purpose of such monitoring lies in providing a stream of indicators that can serve several operational goals:

1. Monitoring during personnel selection activities can provide input for the selection of compatible team members and of individuals with psychological profiles suited to teamwork in extreme environments and situations.
2. Monitoring during training activities can provide diagnostic information useful in guiding further instruction and coaching as well as in determining the composition of teams prior to mission deployment.
3. Monitoring during missions can provide forewarning of potential operational failures due to disruptions of team functioning and give the opportunity to take preventative measures.

These purposes of monitoring make sense only if the collected indicators, whether gathered unobtrusively, through self report, or by other means, are reasonably predictive of outcomes of interest. These outcomes may include subjective and objective assessments of team task performance, team safety performance, accidents, and team-level psychosocial outcomes such as cohesion and morale. In psychometric terms, all indicators obtained from monitoring must be valid assessments of team functioning and must be predictive of some mission outcome of interest.

Unobtrusive monitoring techniques are preferable in the scenarios described above because they would not require the active involvement of personnel in provision of the measures. In addition, given that teams will work in a variety of remote environments, it can be assumed that technology-mediated methods of capturing behavior and communications will be required, because direct observation by supervisors, coaches, or psychologists will generally be feasible only during selection and training activities. With that being said, self-report measures and other assessments that require the active participation of team members may be valuable during a validation phase.

The literature review in this document provides an overview of prior research on the various methods of monitoring personnel performance and the effects that monitoring has on job performance and on other outcomes. Most of this research has arisen from industrial contexts and may not have universal relevance to the space flight context. As a result, we have expanded our view of the literature to include consideration of some areas that have typically not been considered in the research realm of performance monitoring, but may yet provide some worthwhile insights.

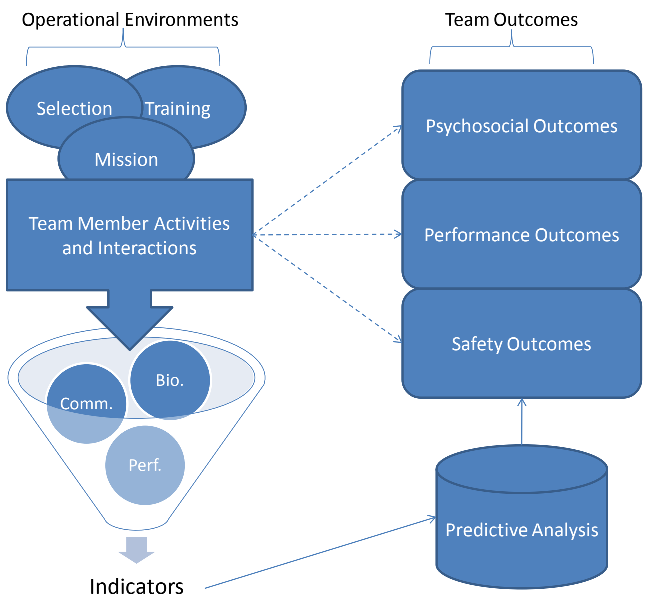


FIGURE 1: CONCEPTUAL OVERVIEW OF THE PROBLEM SPACE

Throughout our analysis of the literature conducted to date, we have assumed that the ultimate goal of this project is to assess the feasibility and options for solutions that would combine unobtrusive collection of indicators with predictive analysis. This assumption is embodied in Figure 1 which reflects our understanding of the problem space. Starting at the top left of Figure 1, we have imagined three operational contexts: a selection context where individual team members are chosen for various roles in a mission; a training context where team members may work together on simulated or practice tasks; and a mission context where the team functions during space flight or in other mission environments. In the latter two contexts, team members interact, perform individual tasks, and collaborate on group tasks. These activities presumably cause the various outcomes experienced during training or missions (three dotted lines pointing right).

We have gathered these outcomes into three gross categories, psychosocial outcomes (e.g., morale, cohesion), performance outcomes (e.g., task completion), and safety outcomes (e.g., mistakes, accidents). We expect that many behaviors and activities reveal observable cues

Formatted: Font: +Headings, 12 pt
Deleted: Figure 1
Deleted: Figure 1
Formatted: Font: +Headings, 12 pt

or indicators about the functioning of a team (the funnel shape on the left). Communicative indicators may include speech and textual communications among team members and between team members and those on the ground. Performance indicators may include intermediate task results and work products (e.g., completion of a subtask in a repair job), physical interactions among team members (e.g., assisting another team member with equipment), or timing indicators (e.g., sleep-wake schedules, time-on task). Finally, another set of indicators focuses on biometrics such as infrared detection of surface blood flow, urinalysis, and galvanic skin response.

To close our consideration of Figure 1, we assume that among the various unobtrusive indicators of individual and team activity, a subset of such indicators may have predictive value in foreshadowing important outcomes such as changes to morale, team performance, or the occurrence of accidents. The cylinder at the lower right of Figure 1 represents an analysis component in which indicators are combined, scored, normed, and compared in an effort to predict outcomes of interest. Throughout the literature review below, we have, in effect, "graded" the existing research with respect to whether we believe it provides promise with respect to indicators, analysis, and/or prediction.

Formatted: Font: 12 pt

Deleted: Figure 1

Deleted: Figure 1

Formatted: Font: 12 pt

4. Industrial Performance Monitoring

To provide a context for the various indicators that could predict team outcomes, we will start by looking at traditional performance monitoring as it has been conducted in industrial settings. In industrial environments such as call centers and manufacturing floors, performance monitoring refers to the gathering of indicators about the work effectiveness and productivity of individuals, groups, and larger organizational units. Prior to the widespread deployment of information and communications technologies, supervisors monitored performance by personally observing, recording, and reporting on employee behavior and work products (Attewell, 1987; Stanton & Julian, 2002). Technological advances over the past 40-50 years such as inexpensive personal computers and networks have facilitated new techniques for performance monitoring and encouraged widespread deployment of these techniques.

Psychologists, sociologists, and others have raised concerns about the use and effects of performance monitoring in the industrial workplace (Stanton & Julian, 2002). In non-military work environments, workers have a range of legal rights – variable across different countries – that influence when, where, and how performance monitoring technologies may be used. Additionally, in many industrial contexts, the existence of labor markets means that work conditions are a source of competitive advantage. As a result of the labor market effects and/or the possibility of employee litigation, many researchers have focused their efforts on understanding how employees *react* to the use of performance monitoring in their work environments. Reactions to performance monitoring have raised sufficient concern that researchers have developed specialized self-report scales specifically for this purpose, such as the one published by Flint (2008). The next section provides a brief overview of this research area.

Reactions to Performance Monitoring

Research on reactions to performance monitoring identifies and explores employees' attitudes and perceptions and relates these concepts to subsequent outcomes. Dependent variables in this research include job attitudes (including fairness, job satisfaction, organizational commitment, and obligation to reciprocate) (Wells, Moorman, & Werner, 2007; Watson, 2008), stress (D. Kiker & M. Kiker, 2008), and mood state (Davidson & Henderson, 2000). For example, the meta-analysis by Kiker and Kiker (2008) showed that electronic performance monitoring was negatively correlated with employee job satisfaction and positively associated with job stress. Researchers have also examined contingent factors that influence reactions and attitudes, such as the existence of feedback (Alder, 2007; Alder & Ambrose, 2005), personality and demographic attributes (J. V. Chen & Ross, 2007), organizational cultures (Alder, 2001), and prior beliefs (Alder, Schminke, Noel, & Kuenzi, 2008).

The issues considered by researchers of industrial monitoring tend to have the greatest relevance in a setting where employees are not used to being monitored, where there are issues of employee retention, and where employees are represented by unions. The space flight context is substantially different in several ways. For example, space flight personnel are monitored frequently on their physical health, are highly familiar with the purposes and goals of self-report measures, and have substantial commitment to activities with demonstrable connections to mission success or safety.

Consistent with Alder (2001), who argues that more "bureaucratic" organizational cultures will respond more favorably to monitoring than supportive cultures, it is reasonable to expect that highly trained space flight personnel may not have the same reactions to performance monitoring as is observed among workers in an industrial environment. Rather than a concern for basic labor rights, space flight personnel may have concerns for the time or inconvenience of monitoring techniques. Space flight personnel also may have concerns around personal privacy, particularly given the confined size of their operational environments.

These concerns suggest that while space flight personnel may have *different* reactions than, say, call center workers, it is no less important to have the cooperation and "buy-in" of space flight personnel with respect to deployed monitoring techniques. An important lesson from the use of monitoring in industrial environments is that employees are creative in finding ways to circumvent controls and mechanisms they consider objectionable (Stanton & Stam, 2006). Therefore, when designing and implementing any monitoring tool for the space flight environment, it will be essential to involve space flight personnel in the processes of evaluating and deploying monitoring tools. Emphasizing to the personnel the advantages of performance monitoring (e.g., safety) and assuring them that they will be protected from the consequences of revealing their mistakes is important as well. Additionally, it is important to be aware of other effects and outcomes that performance monitoring may inadvertently influence. Such effects are discussed in the following section.

Effects and Outcomes of Performance Monitoring

Another question addressed by numerous studies is how performance monitoring affects job/task performance and other related outcomes. Performance refers either to the individual performance within a team or the performance of the whole team, which could be measured with quality of output and quantity of output. Most of the experimental studies in this vein focus on relatively simple clerical tasks such as sorting and editing. Other outcomes refer to safety, errors, and psychosocial aspects that may be applicable to a team, such as team cohesion and morale. Some of the literature in this area focuses on "surveillance," which is a subtype of monitoring used to uncover wrongdoing (D'Urso, 2006). In the industrial context, such monitoring may help to ensure that employees are not stealing, performing sabotage, or procrastinating.

Empirical Research on Effects and Outcomes

Output Quality and Quantity

Quality and quantity of task output are of interest in the present project and therefore it is important to review what previous empirical research has shown when exploring how performance monitoring affects output quality and output quantity. Many studies have focused on how task difficulty influences the quality and the quantity of output given that a performance monitoring system is present.

Davidson & Henderson (2000) found that participants performing an easy task displayed increased task performance under electronic performance monitoring (EPM) and poorer performance when performing a difficult task under EPM. Similarly, Park & Catrambone (2007) sought to investigate whether "virtual humans" embodying the role of the performance monitoring system produced social facilitation effects. They found that for easy tasks, performance in a "virtual human" monitoring condition was better than in the alone condition, and for difficult tasks, performance in the "virtual human" condition was worse than in the alone condition. Consistent with these results from individual studies, in a meta-analysis of EPM literature, Kiker and Kiker (2008) found that EPM has a positive effect on performance *quantity* but a negative effect on performance *quality*. They also ascertained that the EPM-performance quality relationship was moderated by task difficulty such that EPM improved performance quality for simple tasks, but detracted from it for complex tasks. Social facilitation, a theoretical perspective that considers neural system activation and arousal as a basis for changes in performance, is frequently harnessed to explain these effects.

Working in a different theoretical vein, Stanton & Julian (2002) concluded that workers' perceptions of importance of a task were influenced by the capabilities of electronic performance monitoring even though, in all cases, a supervisor stated that both quality and quantity of performance were important. Workers perceived quality to be more important when quality was the only aspect of the task that the system monitored. Workers perceived quality performance to be of lesser importance when only the quantity of performance was monitored. These results suggest that the psychological effects of monitoring with respect to focusing attention and motivating behavior through expectations can be an unintended side effect of both the design of a monitoring system and the communications that managers use to explain and justify the system.

Feedback

Some of the research in industrial monitoring focuses on feedback provided to the employees through monitoring systems. Although performance monitoring has typically been construed as a supervisory activity, the data that monitoring produces can just as easily be used in feedback processes with workers. This line of research generally does not examine the quality-quantity trade-off but rather takes a general view of performance improvement. For example, Alder (2007) found that allowing employees to determine when they receive feedback may enhance their desire to improve. In turn, to the extent that perceptions of interpersonal fairness are high, individuals' desire to respond to feedback will result in improved performance. Similarly, Goomas (2007) discovered that immediate performance feedback and self-monitoring which was delivered to employees improved order picking performance. This improvement was due to an intervention package that included the depiction of goal times and immediate performance feedback.

Conceptual models of Performance Monitoring

The performance monitoring literature contains conceptual models and frameworks guided by psychological theories such as the theory of planned behavior, social facilitation, and the theory of procedural justice. These models portray the relationships between the various factors and outcomes of performance monitoring.

For example, Moran & Nakata (2009) proposed a model based on the Theory of Planned Behavior (Ajzen, 1991) that examines adverse effects caused by ubiquitous monitoring. The theory of planned behavior holds that specific attitudes toward a behavior can predict the occurrence of that behavior. In addition to measuring attitudes toward the behavior, people's subjective norms (their beliefs about how people they care about will view the behavior in question) are also measured (Ajzen, 1991). The factors in Moran & Nakata's model influence factors from the Theory of Planned Behavior and include context, justification (which is affected by trust), awareness, control, boundaries, and intrusion. Behavioral intentions in the Theory of Planned Behavior eventually affect the two outcomes: intended work behavior and unintended work behavior.

Other effects of electronic monitoring are explored by D'Urso (2006) who examined the “panoptic” effects (i.e., a fear of continuous surveillance) of monitoring interpersonal communication in the workplace. According to the model developed in this study, outcomes such as organizational fairness, job performance, workplace satisfaction and others are influenced by organizational management style, organizational communication climate, comfort with technology, and surveillance beliefs.

Cultural dimensions of monitoring were investigated by Panina & Aiello (2005) who proposed a model describing the interaction of major EPM characteristics and national culture dimensions, and suggesting possible implications of this interaction on creating culture-sensitive EPM designs.

The papers reviewed above demonstrate the range of concerns and variables emphasized in performance monitoring research in recent years. It is evident that the focus has been mostly on industrial environments, where the outcomes of interest and the factors influencing these outcomes (such as organizational management style, organizational communication climate, job attitude, fairness perceptions) reflect common characteristics of industrial labor markets. Workers who belong to unions, or who are willing and able to quit a position, or who can raise legal challenges to adverse working conditions have influenced researchers’ decisions about which contexts, variables, and organizations to examine. On a related note, the industrial use of electronic performance monitoring has occurred most frequently in environments where managers are concerned that unmonitored workers may exhibit unproductive or counterproductive behaviors. As a result, the workers and tasks that are monitored tend to be relatively unskilled.

To conclude, although the empirical and conceptual literature on the effects of monitoring is quite thick, much of the literature is only indirectly applicable to the space flight environment. In the literature, performance monitoring is rarely applied to teams and rarely used in the context of highly technical or high level professional jobs. We make a set of baseline assumptions about space flight personnel concerning their levels of organizational commitment, motivation, and task performance that suggest we must look elsewhere in the research for ideas about monitoring the status and functioning of teams.

5. Brief Orientation to Team Dynamics and Performance

Although team dynamics and performance are familiar topics to many readers of this material, we provide a brief overview of them here to uncover a few assumptions that are important to the remainder of this paper. A team is generally defined as a small group of interacting individuals charged with performance of a task, set of tasks, or mission (Guzzo, 1995). The group's membership is well bounded, and members identify with the group.

Research on teams indicates that team performance is a "cross-level" construct, dependent on both the capabilities and characteristics of individual team members and the quality of interaction among them. Interactions among team members fit into a modest number of functional categories: coordination is an important example of a communicative activity that helps keep a team functioning effectively. Leader directives, conflict management, and goal-setting represent other common areas of communication.

These communication processes lead a team through various developmental stages, during which differentiated social and performance roles emerge among team members (Hare, 2003). For example, even in teams without formally assigned leadership, one or more leaders tend to emerge over time. Effective differentiation of roles together with effective enactment of those roles positively influences individual satisfaction, team morale, team cohesion, and team performance.

6. Overview of Possible Indicators

From the brief discussion of teams above, it is evident that indicators of team functioning can be obtained both from individual team members and from interactions among team members. Communications among team members that reveal common understandings of roles, tasks, and goals (as well as areas of dissensus) can provide an important window into how well a team is functioning. Biometric indications of anger and other physiological/emotional states that occur during team interactions, or stress states that persist following team interactions may also provide useful indications of team functioning. Finally, team outcomes, such as intermediate task completion or time-on-task, when compared with established norms or benchmarks, may provide indirect evidence of the quality of team dynamics.

As suggested by Figure 1, a combination of several indicators may help to predict psychosocial, performance, and safety outcomes. In terms of communicative indicators, team mental models, which reflect a shared operational understanding of a mission setting, as well as the status of interpersonal relations on a team, collective beliefs about leadership, and other aspects of team behavior and cognition, may be elicited from textual communications through textual analysis. Along with team mental models, extraction of emotions from text in order to represent the general state of mind of the team and individuals is also a viable option.

Formatted: Font: +Headings, 12 pt

Deleted: Figure 1

Biometric indicators such as keystroke dynamics, facial expressions, gestures, speech, skin temperature, galvanic skin response, and electromyography (muscle activity) may provide another source of information to complete the full picture of how a team of astronauts functions during selection, training or mission. This document will focus mainly on emotion identification at a single moment in time, but a more appropriate usage of a biometric system might track and attempt to identify patterns of change in emotions over time.

Finally, physical interactions among team members (e.g., communicative nonverbal indicators), may be assessed using a strategy known as “proxemics,” an area of research that focuses on the perception, use, and structuring of space. When dealing with proxemics, most often researchers study how spatial use affects and reflects relationships between individuals as members of a dyad or a larger group, and whether the particular use of space is intentional (i.e. seeking interaction) or inadvertent (i.e. in a public setting). In the space flight context, one challenge would be to take advantage of movement and body position in three dimensions. A second challenge lies in the automatic identification and coding of proxemic measures.

Interestingly, some researchers who have been interested in the performance of teams, have approached it from the perspective of underlying cognitive mechanisms instead of overt behavior or motivation. As Rouse et al. (1992) suggested, deficiencies in team coordination, communication and overall performance may be better understood by focusing on underlying mechanisms rather than global behaviors.

One may conceive that the tools for discovering and exposing these underlying mechanisms are a form of performance monitoring, but one that focuses on precursors of complex team activities rather than directly upon the activities themselves. Some researchers who have examined these precursors have focused on “mental models” of complex task performance held by individuals and teams. Research on mental models provides an opportunity to understand how to collect information about a team that may predict later team performance on complex tasks. The next section of the review examines the mental models literature and the techniques used by researchers in this area to extract mental models.

7. Mental Models

Mental Models and Team Mental Models - Definitions

As systems and technologies utilized in the workplace became more complex over the last twenty to thirty years, the issue of individual mental models started gaining interest among researchers. Research had shown that understanding a complex system (e.g., a cockpit) and successfully interacting with it required several different types of knowledge, including knowledge of the basic system components, the possible states of those components and how the components are interrelated (Hegarty, 1991; Rowe & Cooke, 1995). Such knowledge comprises a mental representation, or "mental model," of the system (Gentner

& Stevens, 1983; Rowe & Cooke, 1995; Staggers & Norcio, 1993). A worker who operates complex equipment or system interfaces uses a mental model to understand the systems and any feedback that they provide (Rasmussen & Jensen, 1974; Rowe & Cooke, 1995). Although a well-articulated mental model is not always necessary for effective interactions with complex equipment, mental models are assumed to play an important role in facilitating most human-system interactions, particularly when the equipment behaves in an expected manner (Rowe & Cooke, 1995).

Complex systems often require several operators to work together in order to achieve a goal. One relevant example is the space shuttle's remote manipulator system, which typically requires two coordinated operators for safe and effective use. In such a scenario, each individual needs a well-developed response pattern to external events and the actions of other operators. Thus, shortly after mental models began triggering interest among researchers, *team mental models* also began to gain importance in these research communities. Klimosky and Mohammed's (1994) definition of 'team mental model' asserts that it is an emergent characteristic of the group that is more than just the sum of individual mental models. Although the measurement techniques used to capture team mental models are on the individual level, a team mental model is a group-level phenomenon. As described by the definition, team mental models are team members' shared, organized understanding and mental representation of knowledge or beliefs about key elements of the team's relevant environment.

According to Klimoski and Mohammed, team mental models reflect organized knowledge, internalized beliefs, assumptions, and perceptions. Usually it will be in the form of a set of concepts stored and retrieved from memory in relationship to one another. Such organization may derive from presumed cause and effect linkages, or it may reflect learned patterns. Moreover, while the organized patterns may be "spatial" or "sequential" in nature, most probably, such knowledge is organized semantically. The content of shared mental models might reference representations of tasks, of situations, of response patterns or of working relationships. Allowing for the impact of method and circumstances of measurement, a team mental model represents how the group members as a collectivity think or characterize a set of phenomena associated with effective team performance of complex tasks (Klimoski & Mohammed, 1994). It is possible that multiple mental models (or multiple facets of a single model) coexist simultaneously among team members at a given point in time. These would include models of task/technology, of response routines, of team work, and of social relations (Klimoski & Mohammed, 1994).

With respect to the present review, team mental model research as currently represented in the literature probably has the greatest relevance when imagining a team performance monitoring solution that predicts complex task performance. In contrast, the review below suggests that few if any efforts in the team mental model literature have pertained to the psychosocial status or outcomes of the team. Nonetheless, we present a quite thorough review of the area below in the belief that some of the unobtrusive mental model assessment techniques might eventually be harnessed in support of understanding a broad range of team performance criteria.

Linking Team Mental Models and Team Outputs

Klimoski and Mohammed (1994) argued that the following are important for linking shared mental models with team performance: communication processes, strategy and coordinated use of resources, and interpersonal relations or cooperation. A particular team member must have a conceptualization of what is expected of him or her by each team member for each to jointly succeed (Klimoski & Mohammed, 1994). Team mental models are constructed of both the aggregation of individual mental models regarding the task and the technology, but also of the mental models of how a team operates and what role each team member needs to take.

A great deal of research has been directed by the assertion that since teamwork mental models guide the manner in which individuals perform their tasks and interact with one another, team members who hold *similar* or *aligned* mental models of teamwork are better able to coordinate with one another and thus achieve superior performance outcomes (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). It has been hypothesized that team mental models enable team members to form common expectations, coordinate actions, adapt their behaviors to task demands, facilitate information processing, provide support, and diagnose deficiencies. As such, team mental models influence both team processes (e.g., communication, conflict) and team outputs (Klimoski & Mohammed, 1994; Kraiger & Wenzel, 1997; Marks, Mathieu, & Zaccaro, 2001).

Several studies have been conducted to explore the relationship of team mental models with various team outputs. One of the first on this subject was Rouse et al. (1992) who considered the nature of team performance in complex systems in a context of military training of command and control. Rouse showed that the mental models construct has the potential to provide the basis for a principled explanation of team performance, as well as an avenue for enhancing performance. More specifically, Rouse argued that usage of the mental models construct in terms of the mechanisms underlying the formation of expectations and explanations may enable development of finer-grained understanding of such global team-related phenomena as coordination and communications performance.

Stout et al. (1999) explored the relationship between team planning, shared mental models, and coordinated team decision making and performance in surveillance/defense missions using a commercially available low-fidelity helicopter simulation. Results indicated that effective planning “increased” the shared mental model among team members (indexed as the similarity of individual models to a collective model), allowed them to utilize efficient communication strategies during high-workload conditions, and resulted in improved coordinated team performance.

Mathieu et al. (2000) also used a flight simulator in their study to examine the influence of convergence, or sharedness, of team members' mental models as related to team processes and performance. The general results showed that team processes were related significantly to team performance. More detailed analyses in the same study revealed that team mental model sharedness related significantly to team performance, but the relationship was fully mediated by team processes (e.g., coordination of activities).

Edwards, Day, Arthur, & Bell (2006) used a video game which was designed to simulate a complex and dynamic aviation environment to examine the relationship between the similarity and accuracy of team mental models and compared the extent to which each predicted team performance. The authors presented evidence that, for a task with a defined set of optimal strategies, team mental model accuracy is a stronger predictor of team performance than team mental model similarity. In this case, accuracy was operationalized by comparing trainees' mental models to an expert referent model that served as the "true state of the world." Unlike other research that tends to favor similarity, this pattern of results did not emerge until later in training. In an attempt to explore the determinants of team mental models, this study also provided evidence that team members' ability is related to the development of similar and accurate mental models and that the accuracy of mental models partially mediates the relationship between team ability and team performance.

A study addressing similar constructs by Lim and Klein (2006) examined the relationship between team mental model similarity and accuracy and the performance of combat teams. The teams were expected to perform under high stress and intense time pressure. Their findings suggested that teams whose members organize and structure their team related knowledge in a similar fashion will find it easier to coordinate their activities. These team members are likely to agree upon team priorities and strategies, yielding efficient task performance. Additional findings suggested that team mental model accuracy was also instrumental for team performance. Teams whose average mental models were most similar to experts' mental models performed better than did teams whose average mental models were less similar to experts' mental models.

Team mental model similarity was also explored by Smith-Jentsch et al. (2001), and their findings indicated that higher ranking navy personnel held mental models of teamwork that were more similar to an empirically derived model of expert team performance, than lower ranking personnel. Furthermore, comparisons of mental model similarity within groups of high and low ranking trainees and within groups of high and low experience trainees indicated greater similarity between those of higher rank and between those with greater experience. Another study by the same authors tested the effects of a computer-based training strategy that was designed to develop teamwork mental models that were more similar to the 'expert model' described in the previous study. Using a card sorting approach, positive training effects were demonstrated on similarity to the expert model, similarity to other trainees, and consistency.

As the review above suggests, individual and team mental models have been widely researched in work environments that may have similarities to those where space flight personnel train and work. Moreover, mental models are utilized when the team or individuals need to tackle complex tasks, which is also the more suitable case for our purposes. Finally, the strong (if complex) connection between team mental models and team performance suggests that results from this body of research may be quite relevant to the space flight context. In the next section, we describe the wide variety of possible mental model elicitation techniques. Although many of these techniques are obtrusive and suitable only for research studies, the literature does suggest some possibilities for operational contexts.

Methods to extract and measure team mental models

The means to measure, elicit, or represent mental models in general and team mental models in particular have been discussed extensively in literature on personnel training. Incorporating mental model assessment, diagnosis, and instruction into training requires the selection of an appropriate measure of the knowledge, structure, and assertions in mental models. Because there is no universally agreed-upon measure of this knowledge, selection of a measure can be difficult (Rowe & Cooke, 1995). It has always been challenging to determine the best way to measure mental processes of organized knowledge because these processes are tacit, residing in the person's mind. Therefore, the elicitation of mental models has been a central issue in individual and team mental models research, and various methods have been proposed to extract the information that represents mental processes. Some papers have been written specifically for this purpose while others elaborate on this issue in detail in the methods section due to its importance.

Langan-Fox, Code, & Langfield-Smith (2000) constructed a review describing the potential of each technique for individual and team mental model elicitation and representation. According to the authors, different elicitation techniques require different degrees of researcher involvement, and some techniques are more suited to eliciting an individual mental model than a team mental model. Following are some of the elicitation techniques presented in this review.

Cognitive interviewing techniques:

This category of techniques includes interviews, question-answer interviews, and a technique called inferential flow analysis. A transcript of the interview is constructed and analyzed using propositional or discourse analysis. The final representation is a graph that illustrates domain concepts along with conditional and causal associations among them. Cognitive interviewing techniques can be used to elicit a team mental model directly through group discussion. These group discussions can be used to derive important constructs within a domain and linkages and relationships between those constructs. A disadvantage of group discussion is that, like in any group discussion, often the views of more influential or extraverted group members can dominate the discussion and distort the team mental model in favor of their perspectives. This can be partially overcome by asking each individual to write down his or her responses before group consensus is achieved. An example of application of such technique was an investigation of changes in managers' mental models through the extensive review of questionnaires, interviews and company records (Cavaleri & Sterman, 1997).

VERBAL PROTOCOL ANALYSIS:

This technique is used primarily to obtain information about decision-making strategies and general reasoning processes. It is particularly useful for uncovering decision-making errors attributable to individual biases and misconceptions. Participants are asked to think aloud while they undertake a task or make a decision. Sessions are recorded on audiotape or videotape, and a written protocol is generated afterwards. From the set of recorded verbalizations, the researcher can identify the relationships between objects within a domain. Possible outputs from this technique include sets of production rules, decision trees, heuristics, algorithms, systematic grammar networks, and more. A disadvantage of this technique is that the individual-level output produced by verbal protocol analysis might be difficult to summarize and compare systematically, which limits the usefulness of the technique for team mental model measurement. This technique was applied in the examination of thinking processes in personnel selection (Barber & Roehling, 1993) as well as for physicians' medical reasoning and problem solving (Hassebrock & Prietula, 1992).

VISUAL CARD SORTING TECHNIQUE:

This technique is a quick, easy to administer, flexible, and face-valid way of representing mental models. In visual card sorting, the participant is either provided with researcher-generated concepts or is asked to list all the concepts that he or she sees as relevant to the domain of interest. The concepts are written down on cards, and the participant is asked to sort the cards by placing cards that are perceived to be related closer together. The participant then explains why he or she arranged the cards in such a way. This information is tape recorded or transcribed, and the arrangement of cards (the final representation) is photographed. Although the visual card sorting technique can be used in a group session to measure the team mental model, as with cognitive interviewing techniques, the views of more influential or extraverted group members can dominate the session and distort the model. The use of visual card sorting for team mental model measurement is recommended when research time is limited. One of the studies that used this technique (Daniels, de Chernatony, & Johnson, 1995) examined managers' mental models of competitive industry structures.

Ordered tree technique:

This technique was created as an alternative to multidimensional scaling, when researchers observed that the recall of items in a free recall task often included consistent sequencing of items recalled. While multidimensional scaling suggests that items recalled together may have a "short" distance from each other, the ordered tree technique also considers consistencies in the sequence of items recalled (e.g., when recalling 'a', one tends to next recall 'b' but not vice versa). In this technique, participants are asked to recall a large, well-learned set of items many times from many different starting points, sometimes starting with a cue item and sometimes without. An algorithmic analysis constructs a hierarchical structure among the items based on the resulting sequences. The basic assumption is that respondents have mentally organized items into chunks and will recall the chunks as units, tending to recall a whole chunk before proceeding to the next one. An example of ordered tree technique usage was the investigation of the long term effects of teacher education programs on beginning teachers' cognitive structures for classroom management (Winitzky, Kauchak, & Kelly, 1994).

The ordered tree technique can be used to compare hierarchies between pairs of respondents. Measures of similarity can be calculated between a pair of trees. Perhaps more importantly, team members can discuss the similarities and differences between the hierarchies as a training exercise. The method has often been applied to research that focuses on mental model similarity in expert-novice comparisons.

Causal mapping:

In this technique, the participant is asked whether one concept influences the other, whether it does so positively or negatively, and if it does so weakly, moderately, or strongly for each possible pair of a set of concepts. An $n \times n$ adjacency matrix is then constructed where n is the total number of concepts in the map, and numbers in the cells at the intersection of each column and row indicate the existence, direction, and strength of the relationship between two concepts. A distance ratio formula can be used to infer the extent of difference between the maps of individual team members. An example of such usage of a distance ratio formula is Langfield-Smith (1992) who investigated the collective beliefs about the important aspects of the job of a fire protection officer in a team of firefighters. Markíczy & Goldberg (1995) inspected causal mapping for an individual and proposed a method for expanding causal mapping's value as a tool for exploring individual's idiosyncratic beliefs.

The techniques described above require the presence of a researcher to conduct and guide the process of elicitation. Many of the techniques are time intensive, and in some cases research participants find them annoying. Alternative techniques described below are less obtrusive in the sense that they work from the analysis of incidentally produced materials (such as formal speeches or the recording of the team or individual in action).

Content analysis:

This is a family of systematic methods for analyzing written statements such as formal speeches and transcripts of interviews. The researcher uses a set of coding rules to analyze sentences, phrase by phrase, to uncover important concepts and the relationships between them. Establishing the validity of content analysis for deriving a team mental model is problematic. Content analysis is applied to a corpus of textual data that can be obtained from a variety of sources, such as emails or reports. The purposes of these communications, the circumstances under which they were produced, and the intended audience all influence the information available for analysis. Under optimal circumstances, when a corpus of communications is explicitly focused on the coordination of a team's tasks, it may be possible to derive useful information about individual and team mental models. As a more obtrusive method, interview transcripts can also be used as the corpus for content analysis. One example is the work of Langan-Fox & Tan (1997) who applied content analysis on interview transcripts to investigate organizational culture within a large, government business enterprise.

Observation of task performance:

Researchers can use direct observation of an individual's behavior during the completion of a task to infer mental models. Although complete observation involves a high level of involvement between a researcher and participant, passive observations are also possible. Passive observation entails little or no interaction between the two parties, and the researcher often takes the role of a bystander or uses technological means such as activity logs or videos of task activity to provide indirect evidence of a mental model. A difficulty with passive observation is that it is up to the researcher to identify the important concepts and the relationships between them, and behavior is not always a good guide. For example, a mistake in controlling a system may be due to an inaccurate mental model or simply due to inattention or fatigue. Observation of task performance is best suited to the examination of (individual) mental models in contexts where a user must interact extensively with a system and the sequence of interactions, mistakes, backtracking, and related actions illustrate the nature of the individual's mental model. An example of such work is by Chen (1996) who looked at students' interpersonal cognitive problem-solving skills where behavior observation was one of the methods to elicit those skills. Vandenplas-Holper (1996) used video recordings of children's learning sessions and analyzed using systematic observation to infer changes in mental models over time with increased learning.

Mohammed, Klimoski, & Rentsch (2000) also evaluated a set of techniques for measuring team mental models: Pathfinder, multidimensional scaling, interactively elicited cognitive mapping, and text-based cognitive mapping were critiqued and compared according to their treatment of content and structure, as well as their psychometric properties.

Pathfinder

Pathfinder (PF) is intended to produce psychological scaling of the underlying structure between concepts. The PF algorithm transforms raw, paired comparison data into a network structure in which the concepts are represented as nodes and the relatedness of concepts is represented as links between nodes. Studies that use PF employ an “averaging” technique to transform individual-level data into a team-level cognitive structure. The team level structure can then be compared back to individual structures, and members could then be asked to verify the map for accuracy. It is also feasible for group members to work jointly in order to rate the similarity between constructs and produce a team-level map.

Multidimensional scaling

Multidimensional scaling (MDS) is a psychometric scaling technique that represents proximity data in a spatial map. Given the assumption that geometric distance can represent psychological similarity, MDS can be useful in identifying the unknown underlying dimensions used to cognitively organize stimuli. MDS represents cognitive structures in n-dimensional space. Inputs are most commonly in the form of similarity ratings that respondents provide for pairs of items. The resulting MDS solution, calculated based on similarity data, presents stimuli in relation to the underlying dimensions. Studies that use MDS also average individual-level data to examine team-level cognitive structures. As with PF, no known examples of global measurement exist. However, group members could jointly rate the similarity between constructs to produce a team-level map.

Cognitive mapping Techniques

Cognitive mapping methods are graphic representations of both the content and structure of individuals’ personal belief systems in a particular domain. Cognitive mapping was one of the first cognitive measurement techniques to be introduced into management research and has been used to study decision making, negotiation, organizational cognition, and strategy. Cognitive mapping provides a way of accessing large, untapped sources of data generated by organizations and examines meaning as a relational phenomenon. There are two techniques by which the content to be mapped can be generated. The first, called an interactively elicited causal map (IECM) is obtained by requesting the data from participants through questionnaires and/or interviews and the second, text based causal maps (TBCM), is obtained through post hoc analyses of data (e.g., systematic coding of documents or transcripts).

A similar approach called *map analysis* was employed by Carley (1997), who used this method for extracting, analyzing and combining representations of individual mental models as cognitive maps. This textual analysis technique allows the researcher to extract cognitive maps, locate similarities across maps, and combine maps to generate a team map. Using map analysis, the researcher can address questions about the nature of team mental models and the extent to which sharing is necessary for effective teamwork. Individual cognitive maps can be compared, or combined to create a team cognitive map. If two individual's mental models have been coded as cognitive maps, then these maps can be compared and contrasted. Each individual cognitive map can be thought of as a binary graph (an acyclic, usually tree-shaped structure). As such they can be compared simply by counting the number of shared concepts, shared statements, total concepts, total statements, concepts only in that map, and statements only in that map. Two maps can be combined by creating either a union or intersection file.

Carley (1997) demonstrated this technique using data drawn from a study of software engineering teams. The impact of critical content analysis coding choices on the resultant findings was examined. Various coding choices were shown to have systematic effects on the complexity of the coded maps and their similarity. Consequently, a thorough analysis requires analyzing the data several times under different coding choices. A substantive result reported by Carley is that all teams have comparable models, but successful teams are able to describe their models in more ways than are non-successful teams.

Rowe & Cooke (1995) conducted an empirical study that evaluated four measures to assess individual mental models, with individual task performance as the criterion. The authors tested three methods which involved technicians who tried to deal with a troubleshooting problem: a laddering interview, relatedness ratings, diagramming, and think aloud. Some of these methods are similar to the methods described by Langan-Fox et al. (2000) and Mohammed et al. (2000), but this study also attempted to test these methods' ability to predict performance.

In a *laddering interview*, after being given a troubleshooting problem statement, the technician was asked: (1) to identify the major system important in troubleshooting this problem, (2) to name the major components of the identified system in the context of the troubleshooting problem, and (3) to list all the major components of the identified system, regardless of the problem's context.

For *relatedness ratings*, the technician used a six point scale to rate the functional relatedness of all pairs of the eleven system components. Pairs were presented randomly, and technicians were told to rate them in terms of their first impression of functional relatedness, within the context of the troubleshooting problem.

In the *diagramming task*, the technician arranged and connected index cards, with a component name printed on each, in a manner that represented the way in which the system functions in general. Connections and their directionality were represented with a set of directional and bidirectional arrows.

For the *think aloud* task, the technician stated the troubleshooting actions he or she would take, and a subject matter expert stated the results of those actions. The technician was instructed to verbally express all thoughts, or think aloud, while working to solve the problem.

Of the four techniques assessed, all but the think-aloud technique were predictive of troubleshooting performance. Although the think-aloud verbal reports yielded mental models, these models were not predictive of performance. This may have been because this technique is unstructured, and structuring the think-aloud interview might have resulted in data more closely related to performance. This finding emphasized the importance of verifying that the mental model measure relates to the criterion of interest. The laddering and rating techniques were independently predictive of performance, suggesting that these two measures capture different aspects of a mental model, each of which is important to the troubleshooting task. The laddering task tapped into knowledge about existing components, whereas the ratings task accessed knowledge about the interfaces or connections between components. Both of these measures appeared to be good choices for identifying mental model knowledge when the goal is to improve troubleshooting performance.

While the previous study focused mainly on individual mental models, Smith-Jentsch et al. (2001) reported results from two empirical studies that utilized a card sorting approach to measuring team member mental model similarity in naturalistic training environments. The authors adopted an expert model of teamwork that was derived through the analysis of performance ratings collected from navy command and control teams. This model consisted of four dimensions defined by 11 component behaviors: information exchange (i.e., passing information, providing big picture summaries, seeking information from all available sources), communication (i.e., proper phraseology, brevity, clarity, completeness of standard reports), supporting behavior (i.e., error correction, back-up/assistance), and leadership (i.e., providing guidance, stating priorities).

A card sorting task was used to assess each participant's mental model of teamwork. Each card listed a concrete example of either effective or ineffective teamwork that could occur in a submarine attack center. Participants were instructed to sort the examples into categories of teamwork that were meaningful to them and to label each of their piles. Participants' similarity to one another and to the expert model was computed based on matrices (‘1’ was placed in each cell where the corresponding cards were placed together in a single category) by using the Phi coefficient (the Pearson correlation coefficient between two dichotomous variables). In order to obtain an expert matrix from which to score the accuracy of participants' mental models, three researchers sorted the examples into piles that would be consistent with the expert model of teamwork.

The elicitation strategy presented in Smith-Jentsch et al. (2001) differed from many of those mentioned above in other articles in that most of the team mental model research aspired to extract an aggregated mental model for the whole team, whereas here the idea was to elicit the participants' mental model of *teamwork*, or in other words, what an effective or an ineffective team would look like.

Another empirical study that dealt with measuring team model similarity was performed in the Singapore Armed Forces (Lim & Klein, 2006). Soldiers were randomly assigned to teams, and all teams received the same training program. The soldiers received training in the operations they were about to perform as a team, and also underwent extensive physical fitness training. Data collection took place at two points in time. At Time 1, ten weeks after the teams were formed, the researchers collected survey measures of team members' and subject matter experts' task and teamwork mental models. The task mental model was defined as the team members' shared understanding of the technology and equipment with which they carry out their team tasks as well as their perceptions and understanding of team procedures, strategies, task contingencies, and environmental conditions. The teamwork mental model was defined as the team members' understanding of team members' responsibilities, norms, and interaction patterns together with the team members' understanding of each others' knowledge, skills, attitudes, strengths, and weaknesses.

The authors used their ratings to define the expert (i.e., accurate) mental models. The Time 2 data collection took place 3 weeks following Time 1 data collection, in which the team members' task mental models and teamwork mental models were measured. The researchers asked each team member to judge the relatedness (on a scale 1= unrelated, 7=highly related) of 14 statements describing team procedures, equipment, and tasks. Statements included: 'Team members conducted routine maintenance of their equipment and weapons in the field,' 'Team members are cross-trained to carry out other members' tasks,' 'Team members have a good understanding of the characteristics of the enemy's weapons,' and 'The team is highly effective.' To obtain a measure of each team member's teamwork mental model, the authors asked participants to judge the relatedness (same scale) of 14 statements describing team interaction processes and the characteristics of team members (e.g., 'Team members trust each other,' 'Team members accept decisions made by the leader,' 'Team members communicate openly with each other,' and 'Team members are aware of other team members' abilities'). Once collected, this data was used for measuring similarity between the team members' models and accuracy, as compared to the experts' models.

Team Mental Models: Conclusion

As can be seen from the literature presented above, researchers have explored numerous techniques for team mental model and individual mental model elicitation. Many of these elicitations have led to useful predictions of team performance. Unfortunately these techniques often require intensive researcher involvement in the data collection process and extensive further analysis after data has been collected. These elicitation processes often demand all the team members to be available and fully dedicated to the elicitation tasks. Thus, the prospects for a fully automated and unobtrusive system of mental model extraction during regular mission operations seem limited at this point, although it is certainly possible to imagine a range of future possibilities.

In contrast, the training period on the ground, with more opportunities to gather the team members together without separating one or more of them from their mission tasks, may be a more suitable period for mental model collection than an operational mission phase. In addition, if important disparities between individual or team mental models are identified at the training stage, they could ostensibly be corrected prior to actual negative impacts during operations. Moreover, mental model extraction might usefully guide the trainers in diagnosing deficiencies in a team during the training process.

Assuming that the automation and unobtrusiveness challenges could be met, repeated elicitation of team mental models during mission operations seems to have substantial potential as a method of monitoring a team and predicting future performance. Dissimilarity or lack of accuracy in mental models of the team members can be used as a warning that something in the team may not be functioning properly during the mission. It could indicate a potential for errors, decreased quality or quantity of output, or tensions between the space flight personnel due to misunderstandings.

Although mental models can expose important aspects of team members' thinking, particularly with respect to interactions with complex technologies, as currently construed in the research literature mental models are not useful for predicting all of the outcomes of interest in the present review (e.g., outcomes such as team cohesion or morale).

Other aspects of a team member's state of mind might be obtained by analyzing the communications among the space flight personnel, between the space flight personnel and the mission control, as well as from team members' logs or any other documentation that they are required to provide. In the case of written communications, the text would be directly available for analysis, whereas for oral communications, speech to text would be used. In contrast to the mental model approach, which predicts team performance *indirectly* by examining similarity among the members' models, or the accuracy of models versus an expert criterion, it may be possible to *directly* assess certain variables of interest from the contents of texts. This might allow identification of variables such as cohesion, morale, leadership, or conflicts. In the next section, we discuss a body of literature associated with natural language processing (NLP) or human language technologies (HLT) that contains various methods of extraction of variables of interest from text.

8. EXTRACTING TEAM AND INDIVIDUAL CHARACTERISTICS FROM TEXT

TEXT ANALYSIS OF TEAM MEMBER DISCOURSE

Verbal communication data gathered from members of a team can provide an indication of cognitive processing at both the individual and the team level and can be tied to both the team's and each individual team member's abilities and knowledge (Martin & Foltz, 2004). Team communication provides a source of discourse which can be analyzed and tied to measures of team performance (Foltz, Martin, Abdelali, Rosenstein, & Oberbreckling, 2006).

In one set of studies, team communication processes were hypothesized to mediate the relationship between team member inputs and team performance (Gorman, Foltz, Kiekel, Martin, & Cooke, 2003). This research tested automatic methods that analyzed team communication in order to predict team performance. The text corpora they used consisted of team transcripts that were collected during several experiments that simulated operation of an Uninhabited Air Vehicle (UAV).

In the first study (Kiekel, Cooke, Foltz, Gorman, & Martin, 2002), three different methods were applied: Latent Semantic Analysis, PRONET (Cooke, Neville, & Rowe, 1996), and CHUMS (a method developed for this study). Latent Semantic Analysis (LSA) is a fully automatic corpus-based statistical method for extracting and inferring relations of expected contextual usage of words in discourse (Landauer, Foltz, & Laham, 1998; Martin & Foltz, 2004). This technique can measure the semantic similarity among units of text. Its "knowledge" of the language is based on a semantic model of domain knowledge acquired through "training" on a corpus of domain-relevant text. This training process uses a large corpus of text that has been meticulously tagged by human experts: The software automatically infers the rules used by the human taggers, and these rules can then be used in an automated fashion on future corpora. Because LSA can measure and compare the semantic information in verbal interactions, it can be used to characterize the quality and quantity of information expressed (Martin & Foltz, 2004).

PRONET and CHUMS represent semi-automated analytical strategies that ignore the content of communications but look at the back and forth sequencing of interactions among different speakers. PRONET is a sequential analysis technique that relies on the network modeling tool, Pathfinder (Schvaneveldt, 1990). It is used to determine what events "typically" follow one another, after a given lag. CHUMS is a clustering tool that finds common interaction patterns and then looks for places in the discourse where pattern shifts occur. It works by clustering putative models defined by segments of the sequential data.

Using PRONET, the authors managed to identify variables that can be thought of as a measure of the team's consistency in turn-taking behavior. Turn taking was a useful predictor of performance, primarily during early missions, when skill acquisition was still underway. For CHUMS, the researchers found that measures of team communication consistency are more predictive of performance during the learning acquisition phase of a task. The preliminary results in this study showed strong promise in using automated methods to measure team performance and cognition. Most of the methods were found predictive of performance.

A later study by these authors used LSA as the main method for analysis (Gorman et al., 2003). Within the context of a Predator UAV synthetic task (in which skills pertinent to the corresponding real-world task can be exercised in a controlled setting), the authors developed several methods of communications content assessment based on Latent Semantic Analysis (LSA). These methods include: Communications Density (CD), which is the average task relevance of a team's communications, Lag Coherence (LC), which measures task-relevant topic shifting over UAV missions, and Automatic Tagging (AT), which categorizes team communications. CD and LC were related to UAV team performance. The results showed that the agreement between automatic tagging and a human tagger was comparable to human-human agreement on content coding. The results proved to be promising for the assessment of teams based on LSA applied to communication content.

A subsequent study also applied LSA with the goal to measure free-form verbal interactions among team members (Martin & Foltz, 2004). In this study, the researchers used two approaches to predict the overall team performance scores: by correlating the tag frequencies with the scores and by correlating entire mission transcripts with one another. The results showed that the LSA-predicted team performance scores correlated strongly with the actual team performance measures. This suggests that LSA can be used for tagging content as well as predicting team performance based on team dialogues.

Finally, a more recent study (Foltz et al., 2006) aimed at better understanding and modeling the relationship between team communication and team performance to improve team process, develop collaboration aids, and improve the training of teams. In this study, the researchers used LSA as well, for automating the analysis and annotation of team discourse. Two approaches to modeling team performance were described in this paper.

The first measured the semantic content of a team's dialogue as a whole to predict the team's performance. The second categorized each team member's statements using an established set of discourse tags and used them to predict team performance.

Extraction of Emotions from Text

In order to derive the emotional state of mind of the space flight personnel, as well as the team dynamics among the team members, textual communication may also provide input for automatic text analysis to index the emotional status of an individual or status of relations among two or more individuals. As noted above with respect to latent semantic analysis, the majority of techniques here use human annotated text to "train" the system or to create a model that will be able to recognize these emotions in new unannotated text.

Before creating an automated text analysis system for discovering emotions from text, Rubin, Stanton, & Liddy (2004) tried to answer the question of whether people agree in discerning the types of emotions in text, and if so, to what extent. This paper tied together a theory from social and personality psychology and Natural Language Processing (NLP). The authors presented an empirically verified model of discernable emotions, Watson and Tellegen's Circumplex Theory of Affect, and suggested its usefulness in NLP as a potential model for an automation of an eight-fold categorization of emotions in written English texts. The eight categories that constitute the essence of the theory are: low negative affect (divided to subcategories such as: at rest, calm, placid, relaxed), pleasantness, high positive affect, strong engagement, high negative affect, unpleasantness, low positive affect, and disengagement. Based on the collected data, the authors concluded that the theory is useful as a guide for development of an NLP algorithm for an automated identification and an eight-fold categorization of emotion in texts.

Another study (Mishne, 2005) set out to classify various moods represented in text. Some of the moods are quite similar to the emotions in the Theory of Affect, and some are more "mood like," such as "bored." This study attempted to classify future blog posts by using existing blog text which had been classified according to the mood reported by its author during the writing. That is, given a blog post, the goal was to predict the most likely state of mind in which the post was written: whether the author was depressed, cheerful, bored, and so on. As in the vast majority of text classification tasks, a machine learning approach was applied, meaning that the task was to identify a set of features from the text to be used for the learning process. A variety of features for the classification process were used, including content and non-content features, and some features which are unique to online text such as blogs. The results showed a small, but consistent, improvement over a naive baseline. While the system success rates were relatively low, human performance on this task was not substantially better.

Strapparava & Mihalcea (2008) also utilized blogs and the moods assigned to them. The authors described the construction of a large data set annotated automatically for six basic emotions: anger, disgust, fear, joy, sadness and surprise. The data set consisted of news headlines drawn from major newspapers. The annotators were instructed to select the appropriate emotions for each headline based on the presence of words or phrases with emotional content, as well as the overall feeling invoked by the headline. The annotators used a fine-grained scale, which allowed them to select different degrees of emotional load. For the automatic annotations, the researchers used WordNet Affect, a lexical database where nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept with a subset of synsets suitable to represent affective concepts. In addition to the experiments based on WordNet Affect, the authors also conducted corpus-based experiments relying on blog entries from LiveJournal.com. A variation of Latent Semantic Analysis (LSA) was implemented in this study and compared in performance to UPAR7 which is a rule-based system employing a linguistic approach. The results showed that the UPAR7 system provided the best results of fine-grained evaluations, while the LSA gave the best performance in terms of coarse-grained evaluation.

The approach presented by Francisco & Gervás (2006) also used WordNet for knowledge based expansion of words. This approach considers the representation of emotions as emotional dimensions (valence, arousal and dominance). A corpus of example texts previously annotated by human evaluators was mined for an initial assignment of emotional features to words. This resulted in a List of Emotional Words (LEW) which then becomes a useful resource for later automated mark up. For the actual assignment of emotional features, the proposed algorithm for automated annotation employed a combination of the LEW resource, the ANEW (Affective Norms for English Words) word list (Bradley & Lang, 1999), and WordNet for knowledge-based expansion of words not occurring in either. The method for marking emotions used ideas from two of the main existing methods for marking texts with emotions: keyword spotting and lexical affinity. The algorithm for automated mark up was tested for correctness against texts from the original samples used for feature extraction and against new text samples to test its coverage. Better results were acquired for the texts used to obtain the LEW corpus than for new text.

A list of words or a dictionary was used by Frantova & Bergler as well (2009). This paper explored automatic annotation of dream reports, which were used because they contain information that is mainly not factual, as in newspapers or scientific writing, but rather highly opinionated, sentiment-laden, and emotional. The authors compiled "emotion dictionaries" from a thesaurus using Hall/Van de Castle emotion categories proposed for dream analysis. They managed to capture the inherent ambiguity and polysemy (when a word has multiple meanings) of emotion words in word profiles, which gave a fuzzy membership score of a word on all five emotion categories. The researchers then used the derived dictionaries to assign emotion categories to texts in so-called category profiles. The authors conclude that the system obtained good results when fuzzy category profiles were computed. Fuzziness turned out to be an inherent feature of emotions, but the observed relative ordering and strength encoded in the category profiles seems to be stable even on blog sentence sentiment annotation, a very different text type and task. In general, the comparison with the manual annotation of texts from DreamBank indicated that this multi-faceted approach was promising.

The user's emotional information was used in Guinn & Hubal (2003) to characterize his/her emotional state in interaction with virtual computer characters. This paper describes an effort to develop tagged semantic grammars that carry emotional and attitudinal information about the user's utterance. Semantic grammars are a very common form of language representation for spoken natural language processing systems. These grammars are typically domain dependent, which directly map the incoming text to underlying semantics. In addition to the semantic content of the utterance, emotional and attitudinal information was passed to the dialog manager which utilized this information to modify its model of the user. For example, the designer of the grammar may decide that the use of the word "please" adds to the politeness of the sentence. Thus the rule would indicate that use of the rule in parsing the phrase would increase the overall sentence politeness by a small amount. Values between -1.0 and 1.0 were assigned to emotional tags. Thus a value of 1.0 for POLITENESS would be the maximum value for politeness, while -1.0 would be the most impolite phrase.

Other researchers (Zhe & Boucouvalas, 2002) have attempted to identify emotions through textual interactions such as Internet chat. These authors developed an emotion extraction engine for real time internet text communication that could analyze input text from a chat environment and extract the emotion being communicated as well as the intensity of the emotion. Semantic analysis was used to extract emotional words. Analyzing the individual word position, the person the emotion was referred to, the time the emotion occurred, and identification of emotional words, as well as using a set of grammatical rules allowed the engine to perform satisfactorily. The engine produced better results when analyzing formal writing. Spelling mistakes and slang had a significant negative influence on the engine.

The literature reviewed above demonstrates a variety of techniques based on natural language processing for extracting team and individual characteristics from text. It appears possible to monitor certain cognitive and emotional variables through analysis of textual data generated by space flight personnel (as noted above, speech-to-text would have to be used for oral communications). Although these variables are not team-level constructs such as cohesion and morale, it may be possible to aggregate more basic emotions such as anger, disgust, fear, joy, sadness to infer team level emotions.

The following section contains a consideration of the state of the art in sensing of biological signs, or biometrics. In the context of security and assurance, a variety of sensing devices have been harnessed to assess facial expressions, galvanic skin response, infrared emanations (particularly from the face), and other markers of stress and emotion in individuals. We will be exploring whether biometrics could be used in a time-based, aggregated analysis to reveal some aspects of team functioning.

In addition, a later section focuses on proxemics, or the perception, use, and structuring of space. The research on proxemics studies how spatial use affects and reflects relationships between individuals as a member of a dyad or larger group. In that section, we describe proxemics and examine how it might be incorporated as one of the indicators for team outcomes in the context of space flight monitoring.

9. Biometric Methods

Using and analyzing biometric data may provide another source of information to complete the full picture of how a team of astronauts functions during selection, training or mission. In this review we focus on emotion identification at a single moment in time, but note that the probable usage of a biometric system would be for identifying changes in emotions over time. Commercial products that employ biometric data appear to be in the early stages of development and are utilized mostly for identification and authentication purposes. The academic literature contains a range of ideas regarding the usage of biometric data, and some prototypes exist. Few commercial products exist containing a similar level of capability as these ideas and prototypes.

One type of biometric data to be considered is analysis of keystroke dynamics. For example, Andre and Funk (2005) suggest that biometrics may be used for other purposes than identification of individuals, i.e. to identify individuals' physical health status. These researchers' approach is to detect muscle tension in the users' keyboard usage, to determine the users' individual stress level. In a more recent paper, Vizer et al. (2009) reported on an initial empirical study that investigated the use of timing, keystroke, and linguistic patterns from free text to detect the presence of cognitive or physical stress. Results showed that it is possible to classify cognitive and physical stress conditions relative to non-stress conditions based on keystroke and linguistic features with accuracy rates comparable to those currently obtained using affective computing methods. The proposed approach is attractive because it requires no additional hardware, is unobtrusive, is adaptable to individual users, and is of very low cost. As mentioned above, no available commercial products were found that have a similar functionality, since most of the products that employ keystrokes are designed for security purposes (e.g., for user authentication).

A set of additional biometric measures that have been employed in several studies and prototypes are facial expressions, body movements, gestures, and speech. These indicators can be used to identify emotions either separately or in combination. Based only on speech, Ververidis and Kotropoulos (2006) presented the most frequent acoustic features used for emotional speech recognition and to assess how the emotion affects them and reviewed appropriate techniques in order to classify speech into emotional states. A combination of biometric measures was explored by Busso et al. (2004). They used a recording made by an actress who generated four types of emotions: sadness, anger, happiness, and neutral state. By the use of markers on the actor's face, detailed facial motions were captured with motion capture, in conjunction with simultaneous speech recordings. The results revealed that the system based on facial expression gave better performance than the system based on just acoustic information for the emotions considered. Results also showed the complementarity of the two modalities and that when these two modalities are fused, the performance and the robustness of the emotion recognition system improved. Similarly, Castellano et al. (2008) presented a multimodal approach for the recognition of eight acted emotional states (anger, despair, interest, pleasure, sadness, irritation, joy and pride). This approach integrated information from facial expressions, body movement and gestures, and speech. Fusing the multimodal data resulted in a 10% increase in the recognition rates in comparison with the uni-modal systems.

Skin temperature is another biometric used in several academic articles. Khan et al. (2009), for example, employed facial thermal features in automated facial expression classification and affect recognition. A database of 324 time-sequential, visible-spectrum, and thermal facial images was developed representing different facial expressions from 23 participants in different situations. Another study that used skin related biometric was Nakasone et al. (2005), who described a Bayesian network model that allowed determination of emotion in real time, based on electromyography and galvanic skin response signals. These two signals were chosen for their high reliability. Galvanic skin response is an indicator of skin conductance, and increases linearly with a person's level of overall emotional arousal, while electromyography measures muscle activity and has been shown to correlate with negatively-valenced emotions.

Another biometric measure examined by Kruger and Vollrath (1996) used temporal analysis of speech patterns with a device named LOGOPORT which computes the duration of four parameters for each of the two partners in conversation: (1) Undisturbed speech: when one subject is speaking and the other is listening; (2) simultaneous speech: when one subject is interrupting the other subject and both are speaking simultaneously; (3) pauses in isolation: beginning when one subject stops speaking, and ending when the subject resumes speaking, provided the second subject did not interrupt the first; (4) switching pauses: the time between speaker switching, which means the time that the second speaker needs to take the floor. Note that these measures did not require speech-to-text conversion or any *content* analysis of the actual words spoken. Only the timing and overlap were considered. These speech patterns might be relatively easily obtained in space flight (even for multiple speakers) to gain a deeper understanding of the interactions between the team members.

As mentioned above, utilization of most biometric measures to identify emotion is currently at a research stage, and although this research might eventually be interesting and relevant for the purposes of this project, currently there are no commercial products that are capable of analyzing and utilizing biometric data for the unobtrusive detection of team states.

10. Proxemics

In addition to biometric indicators, we suggest exploration of a set of additional possible potential indicators – related to biometrics because they represent physical cues from a perspective of spatial relationships between people – because they may correspond to the status of relationships in dyads or larger groups. These indicators fall under the category of "proxemics" or the perception, use, and structuring of space. In proxemics research, researchers study how spatial use affects and reflects relationships between individuals as a member of a dyad or larger group, and whether it is intentional (i.e., seeking interaction) or inadvertent (i.e., in a public setting). Notable anthropologist Edward T. Hall was the first researcher who used the term proxemics. Hall developed a notation (coding) system of personal distance based on his extensive observations of humans' use of space and evidence from animal behavior with specific reference to crowding and territoriality. Hall was particularly interested in cultural differences that appeared in people's use of "personal" space.

Methodology in proxemics has focused mainly on interactional settings, for example, how people position themselves in a conversational setting with friends, intimates, or strangers (Harrigan, Rosenthal, & Scherer, 2008). The literature overviewed by Harrigan et al. (2008) shows that the measure coded most often in proxemics is the 'distance' between the interactants. Although it might appear to be a simple task, measuring distance may be not as straightforward as it seems. A variety of different reference points have been used to represent the 'distance' between interactants: measured from their heads, noses, knees, torsos, feet, or chair edges. This issue creates problems in the research literature, because the lack of uniformity and specificity of measurement makes comparing research findings across studies more difficult. Therefore, in some studies, where the independent variable was another interactant's gender, age, culture, or personality (e.g. friendliness, dominance, inconsistency), distance was measured by the seat chosen by the participant or distance she/he approached another participant. One example of such a study is Weitz (1972) who found that participant's chair placement reflected their attitude toward someone of a different race (2008).

While distance is an important variable in proxemics, Harrigan et al. (2008) note that it is a rather limited measure and that Hall's (1963, 1973) approach is more sophisticated and comprehensive. Hall includes the following coding variables: distance, postural identifiers (e.g. sitting, standing), orientation of frontal body plane (i.e. degree one faces another), and input from the senses of touch, vision, audition, olfaction, and temperature (e.g. perceiving heat from another's body). Hall (1963) also divided the spatial world into four social distances, each with a close and far phase, and each based on varying information available from vision, audition, olfaction, thermal reception, and kinesthesia (i.e. sensation of physical alignment of head/body). These four social distances (i.e. intimate, personal, social, and public) span zero to 30 feet, and vary according to the type of interaction and the status of and affiliation between interactants (Harrigan et al., 2008). When it comes to body movement research (kinesics), researchers' coding methods are varied, rarely well defined, and are not often organized conceptually and theoretically. Recently researchers have been attempting to come up with a systematic coding scheme that would apply for both proxemics and kinesics, which include categories such as: trunk lean, trunk orientation, arm positions, leg positions, speech illustrative gestures, self touching, object adaptors, touch, and head actions (nod, shake, tilt, dip, and toss) (Harrigan et al., 2008).

In summary, proxemics, which describes the social aspects of distance between interacting individuals, is another possible indicator to take into account. This distance represents the interactions that occur and provides information valuable to understanding human relationships (Lanz, Brunelli, Chippendale, Voit, & Stiefelhagen, 2009). Proxemic cues of importance for coding interactive behavior include: postural identification (i.e. sitting standing), distance, frontal orientation, and body positioning. Depending on research objectives, touch, eye contact, olfaction, and audition also may be coded (Harrigan et al., 2008).

In the context of this project, measuring the different proxemic variables among space flight personnel might provide indicators regarding their attitudes towards each other. Most of the research reviewed by Harrigan et al. (2008) focused on coding proxemics measures by human coders; however more recent research demonstrated that automatic detection methods that use proxemics and kinesics to detect focus of attention (who is looking at whom), body pose, pointing and hand-raising gestures are also becoming a viable option. In order to gather proxemics data, a wearable device would probably be needed (the use of the Actiwatch[1] for research purposes suggests that such devices might be acceptable to space flight personnel). While proxemics may eventually be capable of automatically providing valuable information that helps in understanding the status and quality of dyadic relationships, there is still research to be done before proxemics could be used as predictive indicators for team performance.

11. OVERALL CONCLUSION OF THE LITERATURE REVIEW

Our review to date suggests that a substantial proportion of the industrial literature on performance monitoring may have limited applicability to space flight personnel, at least in their operational mission environments. There are probably some lessons to be learned about user acceptance of monitoring techniques, privacy concerns, and the influence of monitoring itself on motivation, but the monitoring techniques used in industrial research have been limited mainly to repetitive clerical jobs and tasks requiring a minimum of teamwork.

A promising line of research on the effectiveness of teams seeks to understand and predict team effectiveness through the elicitation of mental models held by team members. In the research, these mental models often pertain to the interaction of a team of users with a complex system. Many of the elicitation techniques described in this research are quite obtrusive and may be unsuitable except in a training environment. Automated extraction of team mental models from communicative texts is a future possibility, but one that is not extensively explored in the current literature.

In contrast, certain variables of interest have been more directly extracted from communicative texts (i.e., without assuming a mental model as an intermediate construct) using automated and semi-automated textual analysis. In one strand of research, communications among team members are analyzed to reveal either patterns of communication (as well as disruption of those patterns) or similarities and differences in the expression of various concepts. In a second strand of research, emotional states or moods have been extracted using machine learning techniques or dictionaries that encode the affective content of various words or phrases. Taken as a whole, these areas of research suggest that monitoring of individuals in teams using natural language processing or spontaneously produced communicative texts may be a viable strategy to pursue.

Finally, we presented biometric and proxemics as areas that may contain an additional set of potential indicators. Literature has shown that utilization of biometric measures to identify emotion is currently at a research stage, and although this research might be interesting and relevant for the purposes of this project, currently there are no commercial products that are capable of analyzing and utilizing biometric data. Most of the research on

[1] http://www.nasa.gov/mission_pages/station/science/experiments/Actiwatch_test6.html

proxemics focuses on manually coding proxemic cues and measures by trained researchers. More recent research demonstrates that automatic detection methods that use proxemics and kinesics may become a viable option. That being said, it is important to point out that there are limitations in the ability of biometric and proxemics measures and detection, specifically the ones that rely on visual input, especially if they will be employed in microgravity environments.

The next two sections will provide an overview of the commercial off-the-shelf products and non-commercial packages that might assist in eliciting mental models and extracting emotions based on textual communication and documents. In addition, the following material also contains summary of interviews of NASA personnel which includes their perspectives on space flight performance monitoring.

12. PRODUCTS OVERVIEW

INTRODUCTION

The literature review above focused on how individual and team performance monitoring can be achieved either with the traditional methods of industrial performance monitoring or with alternative methods that, although not designed to do so initially, may eventually enable unobtrusive acquisition of team mental models and related intrapersonal processes. These methods may provide a window into the individual or shared mental processes. Thus far we've discussed how team mental models may be linked to measuring team outputs and what methods may be used to extract individual and team mental models. We also presented different methods of analyzing text for both mental model elicitation and for extraction of other team and individual characteristics, such as emotions.

At this stage, to provide context for an overview of the commercial and non-commercial products that enable extraction of these characteristics, Figure 2 depicts a working model of several relevant precursors and outcomes. The precursors and variables are based on our interpretation of the specific competencies needed for long-duration missions mentioned in the International Space Station Human Behavior & Performance Competency Model Volume II document (Bessone et al., 2008) and on the dimensions which appeared in the expedition candidate training observation form document (NASA - Mission Operations Directorate Space flight Training Management Office, 2009). Note that this working model is likely incomplete at this stage: a number of other constructs might beneficially be included in a more mature model. Nonetheless, in order to evaluate available text analysis technologies, we considered it important to document our initial thinking.

Table 1 depicts the mapping between the concepts described in NASA documents mentioned above and the terms we use in Figure 2. The model in Figure2 is comprised of three parts: Individual Attributes, Observable Behavior, and Group states. Only the first two parts appear in the table below because only the Individual Attributes and the Observable Behavior factors are competencies, while the third part, Group States, are not competencies, but rather are emergent properties of the group or team.

Table 1: Mapping between NASA-BHP Concepts and Working Model

General BHP Category	Competency and/or Behavioral Markers[2]	Individual Attributes	Observable Behaviors
Self-care self management	"Maintains personal goals in order to feel satisfied and motivated and maximize performance"	Motivation/ Initiative	
	"Refine accuracy of self image; Identifies personal tendencies and their influence on own behavior"	Self Reflection	
Cross cultural	"Demonstrate respect towards other cultures; Understand culture and cultural differences; Build and maintain social and working relationships; Intercultural communication and language skills; Commitment to multicultural work"	Cultural Awareness	
Teamwork and group living	"Acts cooperatively rather than competitively; Takes responsibility for own actions and mistakes; Puts common goals above individual needs; Works with teammates to ensure safety and efficiency; Respects team member's roles, responsibilities, and task allocation"		Active Participation
	"Demonstrates effective teamwork behaviors of performance monitoring, situational awareness, back-up behavior, cooperation, coordination, information & workload sharing"		Coordination and Monitoring
	"Volunteers for routine and unpleasant tasks"		Volunteering
Leadership	"Supports leader; Reacts promptly to Situations requiring immediate response"	Loyalty	

On the left in Figure 2, we depict some Individual Attributes that are precursors to the other constructs in the model. We define these individual attributes as inherent

[2] Based on the International Space Station Human Behavior & Performance Competency Model Volume II and on the dimensions that appeared in the expedition candidate training observation form.

characteristics that each individual brings to the group. As such, these attributes comprise possible components of the astronaut selection process. Next, in the center of Figure 2, Observable Behavior represents the individual and interpersonal activities that occur during the operational mission. We have chosen components that are readily manifested through either verbal or written communication. Communication serves as a window into the behaviors and therefore, we will propose to use primarily textual communication as an input for analysis of the behaviors.

Finally, the Group States are the outcomes that team members experience as a group. As the figure suggests, Group States are the outcomes that the Observable Behaviors cause. The group states are in all likelihood not as readily observable as the communication behaviors but will have to be inferred from other indicators. Of course, a key goal of an automated or semi-automated system in this domain would be to predict adverse changes in one or more of the Group States in advance of their occurrence.

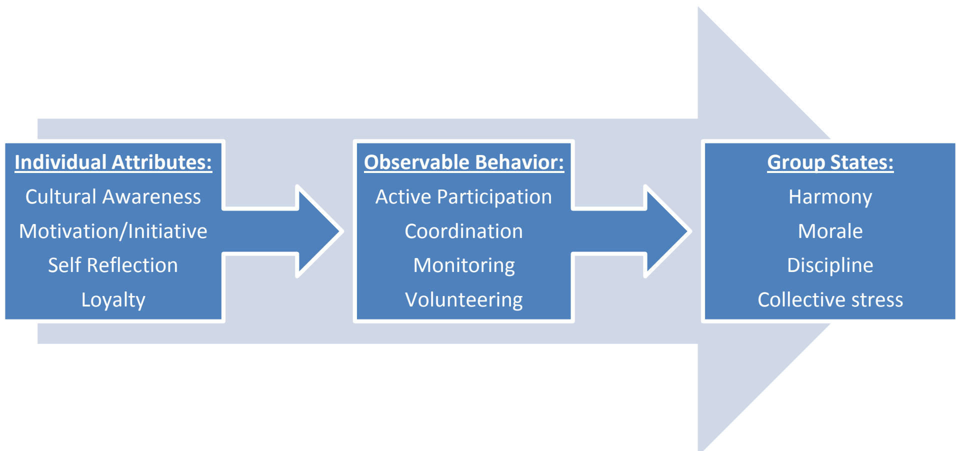


Figure 2: A Working Model of Team Functioning

Focusing on the behavioral descriptions as they appeared in the International Space Station Human Behavior & Performance Competency Model Volume II (and also in the expedition candidate training observation form), we will briefly elaborate on the behaviors to be observed. Active team participation is expressed when a team member acts cooperatively rather than competitively, takes responsibility for her/his own actions and mistakes, puts common goals above individual needs, works with teammates to ensure safety and efficiency, and finally, respects team member's roles, responsibilities, and task allocation. Coordination and Monitoring, on the other hand, occurs when the team member demonstrates effective team work behaviors of performance monitoring, situational awareness, back-up behavior, cooperation, coordination, information sharing, and work load sharing. Finally, Volunteering is when the individual volunteers for routine and unpleasant tasks. Given this breakdown of the behavior, the main questions at this point are whether or not any tools or software packages can be used to identify these behaviors from communicative texts (or speech converted to text) and, once identified, whether they can be used to predict the Group States as they appear in the model above.

Collecting Textual Communication

Communications among team members might provide a window into the state of mind of individuals as well as the status of interpersonal relations among team members. Text analysis methods that produce mental models may in turn provide a deeper understanding of the origin and results of team behaviors. In order to apply text analysis methods, there is a need to gather as much textual communication as possible. Such text-based material could be collected either during training or during a mission. Communications among the team members,either through written messages or text obtained from speech-to-text systems, are most highly relevant, but communications between the team and the ground may also provide useful information. Likewise, individually produced texts (such as logs) might also be informative. All three sources of text may be analyzed either for mental model elicitation or for direct extraction of emotions and behaviors from text. The material below identifies the tools that may enable analysis of text produced by the astronauts.

Comment [k1]: space flight personnel?

A key first step in gathering spoken communication is to perform an automatic transformation from speech to text. The assumption is that recording of either audio only or both video and audio of team interactions will be available. One example of such a package is *VideoLogger* by Virage[3] which uses speech recognition techniques to "watch, listen to, and read" an analog or digital video signal and create a structured video index. *Video Logger* is speaker independent, and it automatically extracts information from video data. The system is intended for semi-automated applications in storyboarding, closed captioning, and related applications. The system can also be configured to recognize faces, voices and types of sounds in the video, identify on-screen text and numbers, and convert spoken words to text. Once a video stream is indexed by *VideoLogger,* the system can be configured to automatically send an e-mail message to designated persons as an instant alert to the existence of specified information.

Dragon AudioMining by Nuance[4] is designed to work on audio only and provides the ability to use text keywords and phrases to automatically search audio files. This software enables the indexing of 100% of the speech information within audio files. By using a speaker-independent dictation engine, it creates XML speech index data for every word spoken within an audio file. The index data includes word, time stamp, confidence levels and metadata associated with the speech information, and can be created from broadcast and telephony-quality sources. A closely related product**,** *Dragon Naturally Speaking,* requires speaker training (i.e., it is speaker dependent) and also requires a special dictionary if unconventional words are used.

Speech-to-text systems have several limitations and challenges. Deng & Huang (2004) found one of the challenges in developing such a system to be the ability to make it robust in noisy acoustic environments. Another challenge to be overcome is the ability to create workable recognition systems for natural, free-style speech (i.e., no pauses between words). In other words, as Deng & Huang noted, the ultimate technical challenge for speech recognition is to make it indistinguishable from the human's speech perception. Shriberg (2005) found some features of natural free-style speech particularly problematic for speech recognition systems such as when people string together sentences without pauses, while on other occasions, people pause (as during hesitations or disfluencies) at locations

[3] http://www.virage.com/rich-media/functions/index.htm
[4] http://www.nuance.com/naturallyspeaking/products/sdk/sdk_audiomining.asp

other than sentence boundaries. According to Shriberg, spontaneous speech has another dimension of difficulty for automatic processing when more than one speaker is involved. An additional challenge area is to "hear" the speaker's emotion or state of being through speech. Modeling emotion and user state is particularly important for certain dialog system applications. We will elaborate more on this type of recognition when we discuss biometric measures in the next chapter.

IBM's *Embedded ViaVoice*[5] is advertised to be able to deal with noise issues and continuous speech. It is available in several languages and provides both speech-recognition and speech-synthesis capabilities. The *Embedded ViaVoice* recognition engine is speaker independent because it is based on small units of speech, called phonemes. According to the developers, the maximum vocabulary supported by *Embedded ViaVoice* exceeds 150,000 words. Based on the known state of the art for other products, however, this package is unlikely to provide accurate, speaker independent recognition of such a large vocabulary.

A recently released commercial tool is Google's automatic captioning for *Youtube*[6] videos. Their system is also designed to deal with free-style speech in the presence of environmental noise, because these are typical characteristics of YouTube videos. No public evaluation results are available for this system, and therefore it is difficult to assess the quality of this recent speech recognition system. Anecdotally, the system provides a minimally useful first approximation of the spoken text that must be edited for accuracy by a human user.

One may bypass the speech-to-text data collection phase by focusing on textual messages created on a keyboard or related input device. Email or computer monitoring software may be used for the purpose of gathering and aggregating all the text that has been typed on a computer. Commercial email monitoring software is intended mostly for surveillance of actions performed by employees on their work computers. Both online and offline activities can be recorded and then reproduced by the employer for viewing and analysis. After reviewing several of these software packages, we found that most email monitoring software functionality is geared toward detailed monitoring of each employee separately. The software does not provide any tools for in depth analysis in order to get the meaning behind the text, but instead it suggests mostly simple presentation of keyword frequencies or alerts on keywords defined by system administrators. In other words, the reports generated by the systems allow only a very shallow form of text analysis. Therefore, this type of software may be useful mostly for text gathering, for further analysis by other tools or software packages.

[5] http://www-01.ibm.com/software/pervasive/embedded_viavoice/about/?S_CMP=wspace

[6] http://www.wiredprnews.com/2010/03/05/youtube-expands-automatic-caption-feature_201003059289.html

One example of such software is *Spector 360*[7]. This package allows very detailed monitoring that enables the employer to see what an employee does each and every second of the work day. It also allows generation of reports and charts across employees to help identify those employees who are most likely engaging in activities that are harmful to the company. A system administrator can define keywords that will be extracted from the text typed by the employee and used to generate charts or alerts that will appear in the report. The main feature relevant to this project's case is the ability to record keystrokes. It includes a keystroke logger that saves keystrokes by application, date and time, and also who typed it, based on user login information. The package also records "hidden" characters and keystroke combinations, such as the Shift and Ctrl key.

Another package that performs keystroke logging is Keystroke Spy[8]. This product also a monitoring solution that can log every keystroke and that captures screenshots of everything they do. It also has an option of delivering alerts when any of a list of keywords is typed. It is capable of logging everything that is typed on the computer or alternatively logging keystrokes typed in specific applications and windows.

Other monitoring tools are capable of alerting not only for specific predefined keywords, but also for predefined patterns. An example of such software is *Mimecast*[9] which includes regular expression testing[10]. This feature's goal is to detect variable data within the text, for example data such as Social Security and credit card numbers or any other data that fits a certain pattern. Although still a very shallow form of text analysis, it could be useful for certain basic detection purposes. For example, there are libraries of words that express extreme affective content (e.g., hate, despise, adore, ecstatic) that could be deployed with this type of software to provide a rudimentary system for flagging messages that express strong emotions.

Once all the communication has been collected, whether it originated in textual form or was transformed to textual form, sophisticated textual analysis may be applied on the text to detect entities, relations, patterns, and other higher order structures. There are no commercial packages that were explicitly designed for the purpose of eliciting mental models, but using text analysis for this purpose or for the purpose of recognizing the existence of a specific behavior might still be feasible, either with a custom developed system or with off-the-shelf systems available in the near future.

---

[7] http://www.spector360.com/

[8] http://www.spytech-web.com

[9] http://www.mimecast.com/email-monitoring/

[10] In computer programming, a "regular expression" is a method for describing a variety of letters, words, or phrases that fit a user-specified pattern. For example, the regular expression "colou?r" could be used to match the alternative spellings color or colour.

TEXT ANALYSIS PACKAGES

In order to examine whether it would be possible to gain insights into individual and team processes in an unobtrusive manner, we reviewed the possibility of applying text analysis to the communications generated by team members in conversation with each other as well as with personnel on the ground. The literature review exposed a family of systematic content analysis methods intended for analyzing written statements such as formal speeches and transcripts of interviews. This is the most unobtrusive method among all the mental model elicitation techniques, and therefore, text analysis software that allows performing content analysis was of particular interest. With proper modeling and "guidance,"[11] this type of software may also be capable of extracting emotions and behaviors based on the text. We looked at several commercial off-the-shelf products that offer textual analysis, usually within an organizational or research context. Below, we present a table summarizing a set of software packages and tools that perform various forms of text analysis, along with their advantages and disadvantages in the context of this project.

Table 2: Advantages and Disadvantages of Text Analysis Software

Name of the software/tool	Advantages	Disadvantages
BusinessObjects Text Analysis	- Extracts information from unstructured text sources such as e-mails, web pages, and documents. - Provides alerts to new or changing information as it develops and allows navigation between relationships, concepts, and timelines.	Intended for general business purposes, requires a human operator to interpret the results and perform further analysis.

[11] In the research area of human language technologies, as well as in related research areas, it is common practice to divide a data set, such as a corpus of text, into a training set and an evaluation set. The training set is used to "train" the software to recognize or detect certain patterns or relationships. Following completion of the training, the software then attempts to recognize the patterns or relationships in the evaluation set. The performance of the software is then gauged by comparison to a known (or derived) statistical benchmark or the results produced by human experts.

Name of the software/tool	Advantages	Disadvantages
PolyAnalyst	- Allows knowledge discovery in large volumes of textual and structured data. - Enables intelligent analysis of data and text by producing easy to understand actionable results. - Allows incorporating dictionaries such as Wordnet.	Operates through interactive drill down and visualizations, all of which require human operators.
TextAnalyst	- Capable of distilling the semantic network of a text completely autonomously, without prior development of a subject-specific dictionary by a human expert. The user does not have to provide any background knowledge of the subject – the system acquires this knowledge automatically.	Operation both on the input side and on the output side is required by a human user.
PASW Text Analytics for Surveys 3.0	- Allows analyzing open ended questions on a survey by employing text analysis and visualization methods. - Quantifying text responses for analysis and automates the process while at the same time enabling to intervene manually in order to refine the results.	Visualization results can be interpreted only by human intervention
Attensity	- Employs semantic approaches to extract and recall information hidden in free-form text, turning it into insights that can be used by all types of business users. - Fusing unstructured and structured data provides an overall picture of the data. - The technology allows users to extract and analyze facts like who, what, where, when and why and then allows users to drill down to understand people, places and events and how they are related.	Intended for business intelligence purposes, the output needs to be manipulated by human users.

Name of the software/tool	Advantages	Disadvantages
Diction 5.0	- This software has a specific purpose of identifying affective tone in a verbal message by performing text analysis.	Only some of the affective tones analyzed by this software are suitable for the purposes of this project.
LIWC 2007	- Capable of detecting emotions and other dimensions in unstructured data.	Word usage in the text needs to be rather explicit for the software to detect the contrast between positive emotions and negative emotions.
KNOT	- This software is built around the Pathfinder network generation algorithm which is a technique to elicit individual and team mental models.	The requirement of the input to be processed and presented in a form of comparison data.

BusinessObjects Text Analysis software[12] by SAP extracts business information from unstructured text sources such as e-mails, Web-based, and customer documents. The vendor suggests using this software to analyze customers, root causes, links, shareholder value, counterterrorism, or employee satisfaction. Main features of this software include entity extraction and analysis, taxonomy-based document categorization, and automatic document summarization. The package can analyze data over time and report on dynamic changes to variables it derives from the data. Once information is collected, the extraction and analysis tools in the software allow navigation of relationships, concepts, and timelines. This software structures language into its most basic parts through automatic language and character encoding identification, document analysis, word segmentation (tokenization), stemming, normalization, decompounding, part-of-speech tagging, and noun phrase extraction.

A similar software package is *PolyAnalyst*[13] by a firm called "Megaputer" (please see the hands-on product review at the close of this chapter). *PolyAnalyst* is a tool for knowledge discovery in large volumes of textual and structured data. This system was designed with the goal to enable firms to answer business questions by scanning unstructured historical data and predicting outcomes of future situations through interactive drill down and visualizations. The interface offers analysis tasks including categorization, clustering, prediction, pattern learning, trends analysis, anomaly detection, link analysis, entity extraction, natural language search, and graphical multidimensional reporting.

[12] http://www.sap.com/solutions/sapbusinessobjects/large/information-management/data-integration/textanalysis/index.epx
[13] http://www.megaputer.com/polyanalyst.php

Another product from Megaputer is *TextAnalyst*[14] which helps users deal with large amounts of text. *TextAnalyst* is intended to summarize, navigate, and cluster documents in a textual database. It can also provide the ability to perform semantic information retrieval or focused text exploration around a certain subject. Specific functionality includes:

- Distilling the meaning of a text - formation and export of a Semantic Network of the text. A Semantic Network is a set of the most important concepts from the text and the relations between these concepts weighted by their relative importance. This network concisely represents the meaning of a text and serves as a basis for all further analysis.
- Summarization of texts – performed by utilization of linguistic and neural network investigation methods. Allows controlling the size of the summary.
- Subject-focused text exploration - user-specified dictionaries of excluded and included words allow the investigation to focus on a chosen subject.
- Navigation through a textual database - the knowledge base can be navigated with hyperlinks from concepts in the Semantic Network to sentences in the documents that contain the considered combination of concepts.
- Explication of the text theme structure - a tree-like topic structure representing the semantics of the investigated texts is automatically developed. The more important subjects are placed closer to the root of a tree.
- Clustering of texts - breaking links representing weak relations in the original Semantic Network enables clustering of the textual database.

In a more research focused domain, *PASW Text Analytics for Surveys 3.0*[15] by SPSS was created for the purpose of analyzing open ended questions on a survey by employing text analysis and visualization methods. Although intended for surveys, it is possible to imagine applying its textual analysis functionality for the purposes of mental model or behavior extraction. The package allows quantifying text responses for analysis and automates the process while at the same time enabling to intervene manually in order to refine the results. The package's main capabilities include identifying major themes, distinguishing between positive and negative phrases, extracting key concepts and opinions, summarizing findings, creating and applying categories, and exporting results for analysis and graphing.

A text analysis package designed for business purposes is *Attensity*[16]. This software has several modules, and the most relevant for this project's purposes are the *Semantic Engines*[17] module and the *Text Analytics*[18] module. The first module employs semantic approaches to extract and recall information in free-form text. The interface allows users to explore the relationships between topics, without having to manually read the whole

[14] http://www.megaputer.com/textanalyst.php
[15] www.spss.com/media/collateral/data-collection/STAS3SPC-0509.pdf
[16] http://www.attensity.com
[17] http://www.attensity.com/en/Technology/Semantic-Engines.html
[18] http://www.attensity.com/en/Technology/Text-Analytics.html

corpus. The software provides keyword search, classification, clustering, categorization, machine learning, case-based reasoning, name entity recognition, language identification, event and relationship extraction, and artificial intelligence. On the linguistics side, it provides exhaustive extraction, advanced pattern recognition, and semantic web.

The *Text Analytics* module automatically extracts data from free-form text. The technology allows users to extract and analyze entities, relations, and events over time. Premade “schemas” are available; these provide aggregated data views which support the schema formats for most of the business intelligence applications in the market.

Most of the tools presented above don’t have a specific analytic goal, but rather are intended to be applicable for a variety of business and research activities. They all assume that a human operator is involved in the process – either on the input side, the creation of the model stage, or on the output side. These packages are incapable of reaching a conclusion or recommending an action; a human user needs to use the output of such software in order to make an informed decision.

In contrast, the following software may be one step closer to one of this project’s goals, which is to extract emotion from text automatically. This software, called *Diction 5.0*[19], is a software package that performs text analysis for the purpose of determining the tone in a verbal message. *Diction 5.0* uses dictionaries (word-lists) to search through text for the following qualities:

- Certainty - Language indicating resoluteness, inflexibility, and completeness and a tendency to speak authoritatively.
- Activity - Language featuring movement, change, the implementation of ideas and the avoidance of inertia.
- Optimism - Language endorsing some person, group, concept or event, or highlighting their positive entailments.
- Realism - Language describing tangible, immediate, recognizable matters that affect people's everyday lives.
- Commonality - Language highlighting the agreed-upon values of a group and rejecting idiosyncratic modes of engagement.

Another software that was designed to detect emotion, as well as other dimensions, is *LIWC*[20] (Linguistic Inquiry and Word Count), which analyzes written or transcribed verbal text files by looking for dictionary terms matched to words in the text. It is done on a word-by-word basis by calculating the percentage of words in the text that match a particular dimension in the dictionary (Sexton & Helmreich, 2003). LIWC includes several dimensions such as linguistic (pronouns, first person, articles, prepositions, etc.), psychological process

[19] http://www.dictionsoftware.com/
[20] http://www.liwc.net/

dimensions (positive emotions, negative emotions, cognitive processes and so on), and more. The software employs dictionaries comprised of words that represent each dimension (for example: the positive emotion dimension is represented by keywords such as happy, pretty, good and other positive terms). It calculates the percentage of words in a section that fall into each dimension.

An additional software program dedicated to a specific text analysis objective is *Knowledge Network Organizing Tool (KNOT)*[21]. One of the methods to extract and measure team and individual mental models is Pathfinder (PF, also mentioned in the literature review), which is intended to produce psychological scaling of the underlying structure between concepts. The PF algorithm transforms raw, paired comparison data into a network structure in which the concepts are represented as nodes, and the relatedness of concepts is represented as links between the nodes. *KNOT* is built around the Pathfinder network generation algorithm. Pathfinder algorithms take estimates of the proximities between pairs of items as an input and define a network representation of the items. The network (a PFNET) consists of the items as nodes and a set of links (which may be either directed or undirected for symmetrical or non-symmetrical proximity estimates) connecting pairs of the nodes. The set of links is determined by patterns of proximities in the data and parameters of Pathfinder algorithms. The system is oriented around producing pictures of the solutions, but representations of networks and other information are also available in the form of structured text files which can be used with other software. The disadvantage of this system is in the requirement of the input to be processed and presented in a form of comparison data. This means that utilizing raw text from communication requires another stage of processing, performed by a human expert.

In summary, the software reviewed above exhibits promising capabilities to transform unstructured text into useful visualizations and other analytic output. These packages provide the opportunity for a human analyst to obtain a sophisticated understanding of a large corpus of text. At this writing, there is no purpose built software that will process a corpus of text obtained from one or more sources and automatically extract from that text a mental model or other high level construct. As can be seen from the disadvantages presented in the table above (Table 2), text analysis may be the closest automatic method to mental model elicitation that is also commercially available. At the same time, this software is still not the perfect tool for unobtrusive acquisition of mental models, emotions or behaviors from text. Text analysis requires a human operator to look at the results, further analyze, and interpret them.

[21] http://pathfindernets.com/KNOT.html

13. A Case Study of *PolyAnalyst* Software

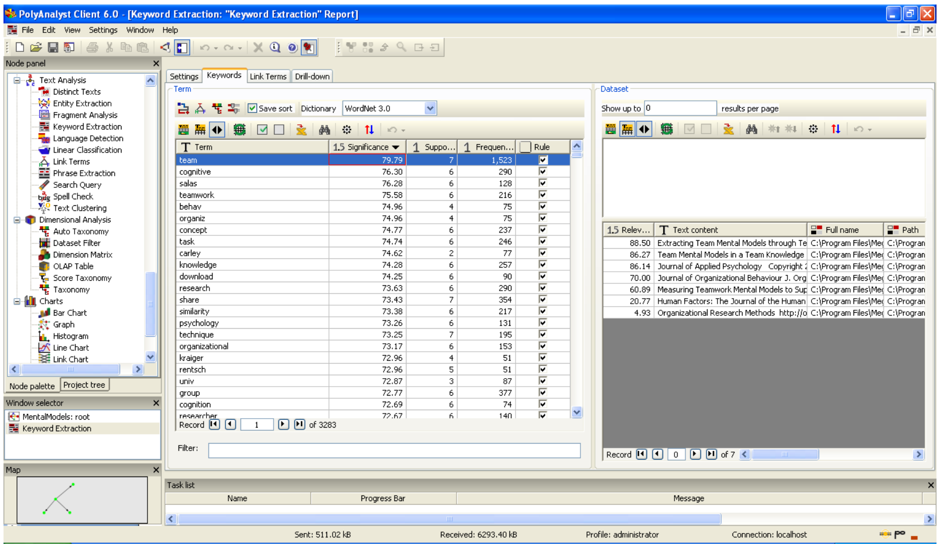
In order to demonstrate how the family of text analysis software packages operates, we performed a very brief exploration of the functionality of *PolyAnalyst*, one of the text analysis packages. This software enables creating a flow of nodes which feed one into another. In other words, the output of one block may be the input of another block, the types of nodes being: data sources, column, row, and table operations, as well as data analysis, text analysis, dimensional analysis, and charts (visualization) functions. The graphical user interface of this software allows using drag-and-drop to choose from the list of available nodes and make the connections between the nodes. Following are a few examples of potential utilization of several *PolyAnalyst* functions.

Given test data in a form of 9 PDF files (containing text from academic papers about mental models) the following functions were used to analyze the text in those files: phrase extraction[22], keyword extraction[23], and auto taxonomy[24]. The primary output of the Keyword Extraction node is a report displaying keywords and information about keywords. For each word in the report, the significance, support, and frequency are listed. The significance is a calculated measure which describes how unique and distinct that keyword is to the current text being analyzed. The support is the number of records which contain the keyword. The frequency is the number of times the keyword appears in all the files. The following screenshot demonstrates how the Keyword Extraction report appears.

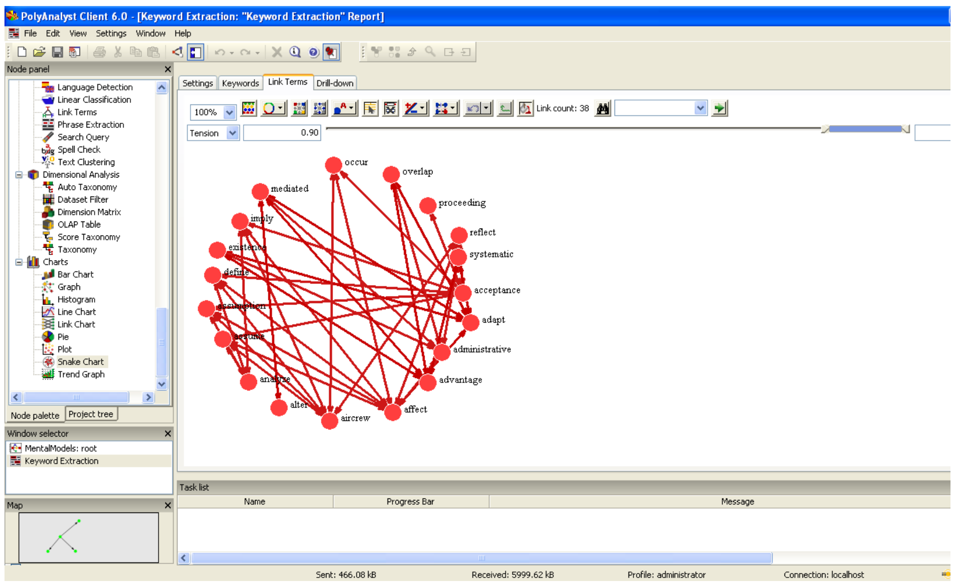
[22] The phrase extraction process derives phrases (a group of alphabetical words which occur next to each other within natural language) statistically by examining the co-occurrences of consecutive words within the text. If two words occur next to each other repeatedly in several sentences across several documents, it can be statistically assumed that these words constitute a phrase.

[23] Keyword extraction derives keywords that are unique and distinct in the current text base being analyzed.

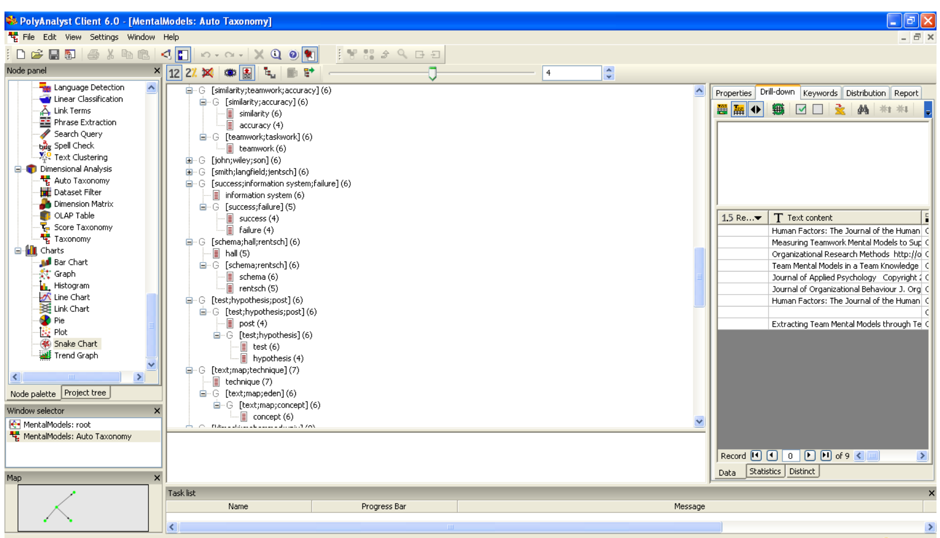
[24] A taxonomy usually has a hierarchical structure similar to a tree. It starts with a root category and underneath the root category there are one or more subcategories (the branches) and possibly more subcategories underneath those. The categories at the very bottom of the tree are often referred to as leaf categories. Auto-taxonomy is derived automatically from the text, based on the co-occurrences of the words.



From the Keyword Extraction report, users can view drilldown results by clicking on a keyword, or drilling down the Link Terms graph and revealing which text contains the selected keyword. The Link Term graph is a visualization that is also part of the Keyword Extraction report. As can be seen in the picture below, this Link Terms option displays a graph of correlated keywords and phrases. By increasing the minimum threshold, the users can filter out relations which have very little support (a low number of records where the two words appear). By decreasing the maximum threshold, the user can filter out some of the very obvious relations (like words which constitute phrases and are always mentioned together).



A feature called Auto Taxonomy presents a taxonomy generated automatically from the text. The tree of categories is displayed on the left, with a count of records matching each category. Similar to the previous functions, clicking on every category in the taxonomy enables drill down to the specific text or records that contributed to the creation of that branch.



As the demonstration suggests, the functionality of this software and other similar packages is usually preset to a limited group of functions. Some of these functions may support the goals of this project to elicit behavior and relations between individuals from text, but this would require turning collected texts into a more structured form.

The output produced requires a human expert to interpret it in order to make an informed decision. Although there are several drawbacks, *PolyAnalyst* does have a few visible advantages that might make this software promising for the purposes of this project. First, among the various methods that can be used for mental model elicitation, the closest method that can be utilized within a software package is content analysis. Instead of using coding rules for content analysis, general patterns can be extracted from the text and keywords are used to represent the text. This software is capable of supporting these functionalities and also enables exploring the links between concepts and keywords within the text. PolyAnalyst allows incorporating various dictionaries (WordNet for example) and machine learning functionalities such as classification, which eventually may lead to the ability to extract emotions and behaviors from text. The workbench style of this software implies that this software enables constructing complex processes that may be operated semi-automatically.

Given the brief exploration above, more information (for example how to transform the text to a more structured form) and further exploration are needed regarding the full potential of this software and its application to the goals of this project. In the past, the potential of this particular software has already been explored by other companies, such as Southwest Airlines, who performed a proof-of-concept demonstration of data and text mining in order to facilitate and promote the use of automated data and text mining for improving overall flight safety performance (Ananyan, Kasprzycki, & Kollepara, 2004). This project proposed new techniques and methodologies to conduct analysis of flight safety data to reveal associations and trends that may otherwise be difficult and time consuming to identify. Although the data used by Southwest Airlines is slightly different and more structured than the data available in this project, parallels may be drawn between the two cases and provide reassurance for further exploring the capabilities of PolyAnalyst for this project. Since other institutions in the aviation community have considered this software, it might have promising potential for this project as well.

14. A Case Study of *LIWC* Software

We chose LIWC software program for this demonstration because it was previously deployed in a study conducted in a setting similar to the one we are investigating. This study, by Sexton and Helmreich (2003), explored the use of language in the cockpit and examined its relationships with workload and performance. The authors chose to study communication within the air crew, because previous research had shown that crew performance was more closely associated with the quality of crew communication than with the technical proficiency of individual pilots or increased psychological arousal as a result of higher workload. Sexton and Helmreich used cockpit communication data that was originally collected for an investigation of the effects of captain personality on crew performance. These data were derived from transcribing four flight segments which involved a three person crew (captains, first officers, and second officers) flying a simulated Boeing 727 during a 5-segment flight over two days. As part of the original data collection, an expert pilot observer was present in the simulator and recorded data regarding individual performance, individual errors, and individual communication skill.

In their study, Sexton and Helmreich found that word count (overall number of words spoken), first person plural ("we"), and number of questions asked in the first flight were positively related to performance and communication as well as negatively related to rates of error. A similar pattern was found for use of the present tense and discrepancy words ("would, should, could"). They also found that captains consistently used more words, used more first person plural, and asked fewer questions than the other crewmembers. Captains also used more present tense than first officers and second officers. The authors presumed that present tense usage is a marker of verbalization such that pilots, who verbalize their actions more, use more present tense, and that linguistic dimension is related to flight outcomes (individual performance, individual errors, and individual communication skill). They also inferred that pilot use of discrepancies could be an indicator of linguistic politeness in the cockpit and that a pattern of increasing use of the first person plural might indicate an increasing sense of familiarity among the crewmembers or an increase in their team perspective.

In order to demonstrate additional capabilities of LIWC, we analyzed two Wikipedia entry discussions. A Wikipedia discussion is a dedicated page assigned to every Wikipedia entry which displays the comments of Wikipedia contributors regarding proposed changes in the contents of that particular entry. The discussion is comprised of comments posted by the contributors in an attempt to settle an issue or disagreement that the contributors have concerning different parts of the entry. The first entry that we chose for this demonstration was a description of a certain event in history, and we analyzed the contributor discussion on how it should be labeled. There was a disagreement among the contributors regarding the labeling of the event because it was politically charged, and therefore a discussion was started on the discussion page. The discussion led to a rather heated exchange of comments fueled by the differences in political perspectives of the contributors. The other Wikipedia entry that we chose to analyze describes a city in New York State, is not charged politically, and has a much "friendlier" discussion around it.

We fed the data from the Wikipedia discussions into LIWC. Each comment was considered as a separate section, and for each section LIWC calculated the percentages (scores) that fell into each dimension. We also edited the dictionary in order to match it to the specific topic and the nature of textual communication. Specifically, “thank you” and “please” were removed from the positive emotion dimension dictionary because these words are usually used as a matter of formal politeness in this type of communication and do not actually reflect positive feelings. In addition, the word “attack” was also removed because it was part of the subject matter being discussed in the comments of the first entry and therefore was used not necessarily to express negative emotions among the contributors. Each comment in the discussion was analyzed separately by LIWC, and thus LIWC calculated a different score for each dimension and each comment. This score reflects the percentage of words in the text that matched the keywords of a specific dimension. For the sake of this demonstration, we looked specifically only at the negemo (negative emotion) and the posemo (positive emotions) dimensions, because they are the most relevant for our study, and we used the scores generated by LIWC to create a box plot.

Box plot[25] is a graphical method for representation of a set of data points. In our case, the value (y axis) is the score that LIWC provided based on the percentages that fell into each dimension. The median of the values is identified by a line inside the box. The body of the box plot consists of a "box," which stretches from the first quartile (the 25th percentile) to the third quartile (the 75th percentile). Two lines (whiskers) extend from the upper edge (top) and the lower edge (bottom) of the box. The upper whisker goes from the top of the box to the largest non-outlier in the data set, and the lower whisker goes from the bottom of the box to the smallest non-outlier. Outliers are marked as small circles on the plot and signify data points that differ greatly from the overall pattern of data.

Our set of data points was based on LIWC output; each data point was a comment in the discussion of the first Wikipedia entry or in the second Wikipedia entry. Figure 3 represents the data points derived from the first entry discussion (on the labeling of a historical event) and Figure 4 represents the data points derived from the second entry discussion (on a city in NY State). As can be seen from the figures below, LIWC was able to identify the differences between the emotions expressed in each of the entries. For both figures, the “0” category represents the negative emotion dimension and the “1” category represents the positive emotion dimension. It can be seen that for the first entry, the one with a more hostile discussion, the percentage of words that express negative emotions (“0”) is generally higher than the percentage of words that express positive emotions (“1”). For the second entry, the one that had a “friendlier” discussion, it can be seen from Figure 4 that the percentage of words that express positive emotion (“1”) is generally higher than the percentage of words that represent the negative emotions (“0”).

Deleted: Figure 3

Formatted: Font: +Headings, 12 pt

Deleted: Figure 4

Formatted: Font: +Headings, 12 pt

Formatted: Font: +Headings, 12 pt

Deleted: Figure 4

[25] http://stattrek.com/AP-Statistics-1/Boxplot.aspx

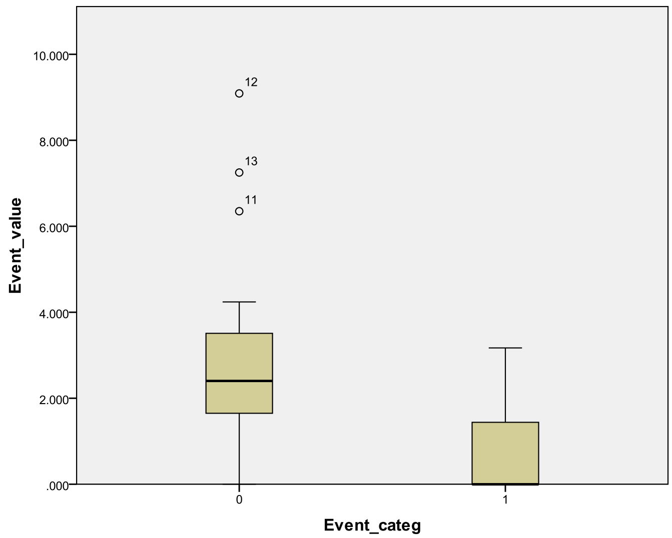


FIGURE 3: BOX PLOT OF LIWC OUTPUT FOR THE HISTORICAL EVENT ENTRY DISCUSSION

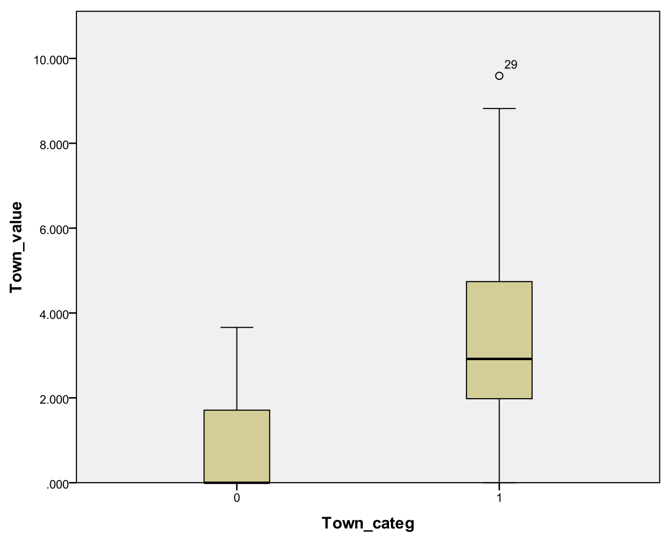


FIGURE 4: BOX PLOT OF LIWC OUTPUT FOR THE DISCUSSION ON THE CITY ENTRY

In summary, LIWC can be a useful tool for easily discovering linguistic and emotional dimensions from transcripts and can potentially add more insight into what takes place in the minds of the team members. The output of the software can be made more accurate if the dictionary is adjusted to the nature of the text being analyzed. It is important to note that based on different attempts and experimentation with various types of text, we have come to the conclusion that in order for the software to be able to discern between the dimensions, especially those associated with emotions, the text needs to contain significant linguistic contrast and explicit wording that is typical to these dimensions.

15. Summary of Interviews with Key Personnel

We interviewed four people who are currently involved in observing or monitoring space flight personnel or are in charge of the technology that allows communication and monitoring. The objective of these interviews was to elicit their perspectives on monitoring team performance during long-duration missions and the feasibility of a potential automatic monitoring system. Our interviewees were: an On-console Communication and Tracking Officer, an Aerospace Psychologist, an Operational Psychologist, and a Flight Controller. Following the constraints of our institutional review board research approval, we withhold the identities of these individuals. For the same reason, we have refrained from presenting verbatim quotes. The following is a synthesis of some of the inputs we got from our respondents.

Currently, 24 hour video transmission is not provided from the International Space Station. Availability of video depends on permission given by the astronauts. For public affairs events, HD video and audio are transmitted to the ground, but for all other purposes, the video is in standard definition, and the level of audio quality is variable. Noise on board might be an issue with respect to picking up dialog among the astronauts, and monitoring technology will need to compensate for that. Besides the traditional means of communication with the ground, the astronauts also have email and twitter access. A suggestion of using gyroscopes in order to identify human movement on the station would likely be superseded by more direct means of assessing personnel activity, such as the "Actiwatch" mentioned previously in this review.

While there are no joint international operations, personnel often reside in their own modules and have lesser levels of interaction with their international colleagues. This might be due to the fact that sometimes the astronauts go on a mission after they had seen the other members of the crew (especially the international ones) only a couple of times. Retreating to their modules may also indicate social frustrations among the astronauts. Observing the dynamics between the astronauts from different cultures might be of value because there are some culture specific traits that may lead to conflicts (in spite of extensive training). Some ground personnel have noticed a "curve" of progress in social relations, where the astronauts' liking increases as time progresses, then levels off, and then improves again near the end of their mission period.

When asked about indicators that would suggest problems in team functioning, our interviewees mentioned fatigue, which may be expressed by limited conversation among the team members. On the other hand, a lively discussion that turns into a heated argument may also be an indicator of problematic team functioning. Usually, when things are going poorly, the crew will bring it up in one of the private sessions that the astronauts have with their psychologists and complain about irritation, not getting along with someone, personality clashes or team dysfunctions.

In terms of assessing team performance, the interviewees suggested examining the type of interactions, inquiries, questions, asking for assistance, how many declarative statements are made, how many coordinating statements are made, how many interactions occur per minute, how quickly a team gets the task completed, and instances where the astronauts are not following procedures. Some crews on shuttle flights have higher error rates than others. A number of tasks are more critical than others, and the tasks that require more focus and might be dangerous are: launch, docking, ISS-Soyuz relocation, robotic arm operations, EVA (extra vehicular activities) coordination with crew inside and outside, and landing.

Sometimes mission control personnel have observed the dynamics between the astronauts in body language, posture, and distance between the crew members. Ground personnel try to detect deviations from the crew's usual behavior. They pay attention to open air to ground channels, such as when sarcastic remarks about the space or ground crew are couched in humor. Some tensions may be evident from the videos transmitted to the ground, for example an incident of microphone grabbing between the crew members during a public affairs event. One of the interviewees noted that the best way to predict problematic team performance so far has been to gather feedback from the crew itself through verbal self report. This respondent incorporated a self report system with his ground crew and inferred from this process that as long as personnel feel protected from the upper management, they do not hesitate to report their mistakes and suggest what can be improved.

We asked interviewees about reactions to monitoring as well. We learned that one of the reasons why video is at the astronauts' discretion is because they don't want to feel critiqued. Also, when the cameras are on, the astronauts will try to present themselves as likeable and task oriented, because they want to be assigned to more missions. Historically, in earlier phases of space activity, performance led to rewards or punishment, and this has resulted in a reluctance to be monitored. Astronauts want to come across as very confident and not to show that they might be having difficulties. In order to improve astronauts' reaction to monitoring, astronauts must buy into the importance of the monitoring system and be shown that its advantages are significant, such as reducing errors, increasing time efficiency or safety.

Overall, the interviews provided us with insights regarding the current state of monitoring the astronauts and their potential reactions to automatic monitoring, as well as what we should take into consideration or pay more attention to when designing a monitoring system. We learned that negative reactions to monitoring may exist, and that there is a need to convince the astronauts of benefits from monitoring as well as to ensure that they will be protected from adverse uses of monitoring data. In addition, we confirmed that other aspects such as body language, proximity between the astronauts especially from different cultures, and deviations from their normal behavior could be considered as indicators.

16. Overall Conclusions

The literature review portion of this document showed that the research on industrial performance monitoring has limited value to space flight operational mission settings. The review suggested that a more relevant line of research exists focusing on the effectiveness of teams and how team effectiveness may be predicted through the elicitation of individual and team mental models. Note that the "mental models" referred to in this literature typically center on a shared operational understanding of a problem space, such as the cockpit controls and navigational indicators on a flight deck. In principle, however, it is not difficult to imagine that such mental models exist reflecting the status of interpersonal relations on a team, collective beliefs about leadership, success in coordination, and other aspects of team behavior and cognition. Although many of the elicitation techniques described in the literature review are quite obtrusive, and may be unsuitable except in a training environment, the second part of this document provided an overview of the available off-the-shelf products that might reflect future possibilities for extraction of mental models and elicitation of emotions based on the analysis of communicative texts. Another possibility explored in this document is the option of incorporating biometric measures in order to expose various individual states (such as stress) that may be indicators or predictors of certain elements of team functioning.

The search for text analysis software or tools revealed that currently there are no available commercial off-the-shelf tools that enable extraction of mental models automatically and unobtrusively, relying only on collected communication text. Commercial text analysis software is, on the one hand, too general and, on the other, not flexible enough to be operated without human intervention. Therefore, usage of this software to derive how a team is functioning and what its mental models are may be relevant for the selection or training stages, when human operators are available. Alternatively, if output from the software described above or from a modified version can be sent to the ground periodically and analyzed by experts on the ground, then these software packages might be employed during missions as well. Clearly, since the packages and tools reviewed in this document were designed mostly for business purposes, utilizing them as-is will not be optimal, and adaptations to the space flight context will be required. Nevertheless, the core capabilities of these packages may be useful as a starting point.

In addition, emotion detection software applications may be useful tools to easily discover linguistic and emotional dimensions from transcripts and potentially add more insight into what takes place in the minds of the team members. The disadvantage in this type of software is that in order for the software to be able to discern between the dimensions, especially the ones associated with emotions, the text needs to contain significant linguistic contrast and explicit wording that is typical to these dimensions.

Biometric and proxemics comprise a variety of indicators which have their own limitations and have not been incorporated into any known off-the-shelf commercial software packages. The advantage of these methods is that, unlike text based indicators, these indicators rely on less explicit cues which may not be expressed through text. The disadvantage lies in the fact that more research may be needed in these areas in order to adjust these indicators to the space flight environment and interaction.

The interviews we conducted with personnel currently involved in observing or monitoring astronauts helped us obtain their perspectives on monitoring team performance during long-duration missions and the feasibility of a potential automatic non-obtrusive monitoring system. Their input suggests that negative reactions to monitoring may occur, and that there is a need to convince the astronauts of the importance and benefits of automatic monitoring.

Together, the literature and evidence we reviewed suggest that unobtrusive monitoring of space flight personnel is likely to be a valuable tool for assessing team functioning in future missions. Similar to results from research on electronic monitoring in industrial environments, it is important to have “buy-in” from the personnel who are affected by such monitoring. Certainly, keeping monitoring unobtrusive will help with this process, but the uses and outcomes of monitoring are important dimensions influencing acceptance as well. Several research gaps must be filled in our understanding of what indicators to collect and what analyses to apply before prototype systems can be developed that will provide data about team effectiveness.

17. Future Steps and Research Recommendations

The literature review and operational assessment presented in this document described some of the directions one might pursue in order to design and eventually create systems that would enable monitoring team outcomes based on various indicators. Some of the indicators mentioned above will require more research and adaptation than others. When considering these options, note that the composition of the “team” under study might be considered not only as the space flight crew itself, but also as a larger collective that includes the ground control personnel. This is a particularly important consideration given data from the interviews: ground crew members may be able to contribute perceptual or objective "criterion" data about the status and performance of space flight teams. These or other criterion data will be required to assess the usefulness of the various indicators being examined. Although it was beyond the scope of the current review, it would be valuable for future research to develop a "directory" of available space flight team performance criteria (e.g., time on task, task error rates) for use in validation studies. With the present state of the art, even if we had the means of collecting reliable predictors of team performance, we might have difficulty validating them for lack of systematically collected criteria.

Although biometrics and proxemics are interesting and promising areas, they still require considerable additional research and evaluation before a workable unobtrusive monitoring system could be designed and implemented. Such an effort would probably need to begin with human review and coding of videotape or other data streams in an effort to observe patterns with relevance to team performance. Given a preliminary understanding of those patterns, both hardware and software prototypes would be needed in order to perform a proof of concept for measures gathered from biometric traces or proxemic cues. A combination of several types of data such as facial expressions, gestures, speech, skin temperature, and proxemic cues might be able to provide a relatively complete picture of team interactions and functioning.

These types of data could preferably be gathered by a small, wearable data collection device, such as the Actiwatch. For example, if the Actiwatch or a similar product could be enhanced to record the proximity between two individuals, measure skin temperature, and record the speech timing of each team member, this might provide a rich source of data for later analysis. A device similar to LOGOPORT (Krüger & Vollrath, 1996), that analyzes speech patterns, could be either part of the device itself or could collect recordings for later analysis. Much of this work could be piloted in analog environments with the beneficial side effect that large data sets of sensor data might serve as a resource for additional research.

Compared with biometrics and proxemics, textual analysis is a more mature and established research area. Notably, open source and commercial software packages are readily available and capable of performing semi-automatic analysis on large text corpora. The disadvantage of textual analysis software packages is that they are currently not adjusted to the requirements of extracting team mental models and often require setup and interpretation by a human operator.

In order to perform further research and further develop existing text analysis software, there is a crucial need to obtain large corpora of actual communication data such as transcripts of communication among team members, discussions with ground control personnel, mission logs, and astronauts' personal journals. Thus, an important emerging research need is the collection, transcription, and annotation of “natural” texts that spring from interactions among team members either on the International Space Station or in one or more of the analog environments (e.g., NEEMO) where research is conducted. In fact, an initial step in this area would be to conduct a review and feasibility analysis of the various possible sources of text throughout the space flight research ecosystem (including operations in all of the analog environments as well as archival recordings of earlier missions). One valuable goal of such a review would be the development of plans for a data repository, where reusable data, scrubbed to varying levels of anonymity, would become available for use in subsequent research projects.

Mental model elicitation techniques that are currently performed manually by a human operator will need to be translated into software modules or algorithms that will be capable of automatic analysis. If machine learning will be employed, a corpus developed from transcripts and communication texts would be used to train the algorithms to deal with the type of text that represents the space flight domain and terminology. To this end, it will be necessary to develop new dictionaries for use with tools such as LIWC and new workflows for products such as PolyAnalyst. Once the software is enhanced or implemented, the predictive power of communication-related indicators must be confirmed by testing the software, together with a new text corpus and the relevant criterion data, preferably from a current operational environment or analog.

As previously described, it will be essential to involve space flight personnel in the processes of designing, evaluating, and deploying any future monitoring tools. As soon as a promising area of investigation or a candidate technology is selected, further efforts to obtain reaction data from subject matter experts (e.g., space flight personnel who have recently completed one or more missions) using mock-ups of systems and results will help to ensure that a subsequent validation effort or other deployment of a working system will proceed smoothly. Judging by the insights provided from interviewees, the organizational and cultural issues existing among managers, space flight trainees, ground crews, and others may provide substantial barriers to successful implementation, even of a technically sophisticated and effective monitoring system. Thus, rather than focusing exclusively on developing the operational capabilities of a technical solution for unobtrusive team monitoring, we suggest a parallel and simultaneous focus on the “contextual” issues that may enhance or inhibit the successful deployment of tools that can predict space flight team effectiveness. Understanding how to overcome the organizational culture barriers to the deployment of an unobtrusive monitoring system may in the end have equal importance with the technological quality of the system.

18. References

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179–211.

Alder, G., Schminke, M., Noel, T., & Kuenzi, M. (2008). Employee Reactions to Internet Monitoring: The Moderating Role of Ethical Orientation. *Journal of Business Ethics*, *80*(3), 481-498.

Alder, G. S. (2001). Employee reactions to electronic performance monitoring: A consequence of organizational culture. *The Journal of High Technology Management Research*, *12*(2), 323-342.

Alder, G. S. (2007). Examining the relationship between feedback and performance in a monitored environment: A clarification and extension of feedback intervention theory. *The Journal of High Technology Management Research*, *17*(2), 157-174.

Alder, G. S., & Ambrose, M. L. (2005). An examination of the effect of computerized performance monitoring feedback on monitoring fairness, performance, and satisfaction. *Organizational Behavior and Human Decision Processes*, *97*(2), 161-177.

Ananyan, S., Kasprzycki, R., & Kollepara, V. (2004). *Application of PolyAnalyst to Flight safety Data at Southwest Airlines*. Megaputer Intelligence.

André, J., & Funk, P. (2005). A Case Based Approach Using Behavioural Biometrics to Determine a User's Stress Level. In *Workshop proceedings of the 6th International Conference on Case Based Reasoning. Chicago, editor (s): Isabelle Bichindaritz, Cindy Marling* (p. 9).

Attewell, P. (1987). Big brother and the sweatshop: Computer surveillance in the automated office. *Sociological Theory*, *5*, 87–99.

Barber, A. E., & Roehling, M. V. (1993). Job postings and the decision to interview: A verbal protocol analysis. *Journal of Applied Psychology, 78*, 845–845.

Bessone, L., Coffey, E., Filippova, N., Greenberg, E., Inoue, N., Gittens, M., Mukai, C., et al. (2008). *International Space Station Human Behavior & Performance Competency Model - Volume II*.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings.* (Technical report). The Center for Research in Psychophysiology, University of Florida.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., et al. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces* (pp. 205–211).

Carley, K. M. (1997). Extracting Team Mental Models through Textual Analysis. *Journal of Organizational Behavior, 18*, 533-558.

Castellano, G., Kessous, L., & Caridakis, G. (2008). Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech. In *Affect and Emotion in Human-Computer Interaction* (pp. 92-103).

Cavaleri, S., & Sterman, J. D. (1997). Towards evaluation of systems thinking interventions: a case study. *System Dynamics Review*, *13*(2), 171–186.

Chen, J. V., & Ross, W. H. (2007). Individual differences and electronic monitoring at work. *Information, Communication and Society, 10*, 488-505.

Chen, Y. (1996). An experimental study of the influences of cognition of interpersonal problems and problem-solving. *Psychological Science China*, *19*, 282–286.

Cooke, N. J., Neville, K. J., & Rowe, A. L. (1996). Procedural network representations of sequential data. *Human-Computer Interaction*, *11*(1), 29–68.

Daniels, K., de Chernatony, L., & Johnson, G. (1995). Validating a method for mapping managers' mental models of competitive industry structures. *Human Relations*, *48*(9), 975.

Davidson, R., & Henderson, R. (2000). Electronic Performance Monitoring: A Laboratory Investigation of the Influence of Monitoring and Difficulty on Task Performance, Mood State, and Self-Reported Stress Levels. *Journal of Applied Social Psychology*, *30*(5), 906-920.

Deng, L., & Huang, X. (2004). Challenges in adopting speech recognition. *Communications of the ACM*, *47*(1), 69-75.

D'Urso, S. C. (2006). Who's watching us at work? Toward a structural-perceptual model of electronic monitoring and surveillance in organizations. *Communication Theory*, *16*(3), 281.

Edwards, B. D., Day, E. A., Arthur, W., & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology*, *91*(3), 727-736.

Flint, D., Haley, L., & McNally, J. (2008). Development of a Scale to Measure Reactions to Electronic Monitoring. *The Business Review, Cambridge*, *11*(2), 131.

Foltz, P. W., Martin, M. J., Abdelali, A., Rosenstein, M. B., & Oberbreckling, R. J. (2006). Automated team discourse modeling: Test of performance and generalization. In *Proceedings of the 28th Annual Cognitive Science Conference* (pp. 1317–1322).

Francisco, V., & Gervás, P. (2006). Exploring the compositionality of emotions in text: Word emotions, sentence emotions and automated tagging. In *Proceedings of the AAAI-06*

Workshop on Computational Aesthetics: AI Approaches to Beauty and Happiness, Boston.

Frantova, E., & Bergler, S. (2009). Automatic Emotion Annotation of Dream Diaries. Presented at the K-CAP Workshop on Analyzing Social Media to Represent Collective Knowledge.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Routledge.

Goomas, D. T. (2007). Electronic performance self-monitoring and engineered labor standards for "man-up" drivers in a distribution center. *Journal of Business and Psychology*, *21*(4), 541–558.

Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. J., & Cooke, N. J. (2003). Evaluation of Latent Semantic Analysis-Based Measures of Team Communications Content. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 47*, 424-428.

Guinn, C., & Hubal, R. (2003). Extracting emotional information from the text of spoken dialog. In *Proceedings of the 9th International Conference on User Modeling.*

Guzzo, R. (1995). Introduction: At the intersection of team effectiveness and decision making. *Team effectiveness and decision making in organizations,* (p. 1-8), New York: Sage Publications.

Hall, E. T. (1963). A System for the Notation of Proxemic Behavior. *American Anthropologist*, New Series, *65*(5), 1003-1026.

Hall, E. T. (1973). *Handbook for proxemic research*. Washington, DC: Society for the Anthropology of Visual Communication.

Hare, A. (2003). Roles, relationships, and groups in organizations: Some conclusions and recommendations. *Small Group Research, 34*(2), 123.

Harrigan, J., Rosenthal, R., & Scherer, K. (2008). *New Handbook of Methods in Nonverbal Behavior Research* (1st ed.). Oxford University Press, USA.

Hassebrock, F., & Prietula, M. J. (1992). A protocol-based coding scheme for the analysis of medical reasoning. *International Journal of Man-Machine Studies*, *37*(5), 613–652.

Hegarty, M. (1991). Knowledge and processes in mechanical problem solving. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving* (pp. 253-285). Routledge.

Khan, M. M., Ward, R. D., & Ingleby, M. (2009). Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature. *ACM Transactions on Applied Perception (TAP)*, *6*(1), 6.

Kiekel, P. A., Cooke, N. J., Foltz, P. W., Gorman, J. C., & Martin, M. J. (2002). Some Promising Results of Communication-Based Automatic Measures of Team Cognition. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 46,* 298-302.

Kiker, D., & Kiker, M. (2008). A Quantitative Review of Organizational Outcomes Related to Electronic Performance Monitoring. *The Business Review, Cambridge*, *11*(1), 295.

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

Kraiger, K., & Wenzel, L. H. (1997). Conceptual development and empirical evaluation of measures of shared mental models as indicators of team effectiveness. *Team performance assessment and measurement: Theory, methods, and applications*, 63–84.

Krüger, H. P., & Vollrath, M. (1996). Temporal analysis of speech patterns in the real world using the LOGOPORT. In *Ambulatory Assessment. Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies* (pp. 103–113). Seattle: Hogrefe & Huber.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, *25*, 259–284.

Langan-Fox, J., Code, S., & Langfield-Smith, K. (2000). Team Mental Models: Techniques, Methods, and Analytic Approaches. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *42*(2), 242-271. doi:10.1518/001872000779656534

Langan-Fox, J., & Tan, P. (1997). Images of a culture in transition: Personal constructs of organizational stability and change. *Journal of Occupational and Organizational Psychology*, *70*, 273–294.

Langfield-Smith, K. (1992). Exploring the need for a shared cognitive map. *Journal of management studies*, *29*(3), 349–368.

Lanz, O., Brunelli, R., Chippendale, P., Voit, M., & Stiefelhagen, R. (2009). Extracting Interaction Cues: Focus of Attention, Body Pose, and Gestures. In A. Waibel (Ed.), *Computers in the Human Interaction Loop* (pp. 87-93).

Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403-408.

Markiczy, L., & Goldberg, J. (1995). A method for eliciting and comparing causal maps. *Journal of Management*, *21*(2), 305-333.

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

Martin, M. J., & Foltz, P. W. (2004). Automated team discourse annotation and performance prediction using LSA. In *Proceedings of HLT-NAACL 2004: Short Papers on XX* (pp. 97-100). Boston, Massachusetts: Association for Computational Linguistics.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273-283.

Mishne, G. (2005). Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access.*

Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The Measurement of Team Mental Models: We Have No Shared Schema. *Organizational Research Methods*, *3*(2), 123-165.

Moran, S., & Nakata, K. (2009). Ubiquitous Monitoring and Human Behaviour in Intelligent Pervasive Spaces. In *Computational Science and Engineering, IEEE International Conference on* (Vol. 4, pp. 1082-1087). Los Alamitos, CA, USA: IEEE Computer Society.

Nakasone, A., Prendinger, H., & Ishizuka, M. (2005). Emotion recognition from electromyography and skin conductance. In *Proc. of the 5th International Workshop on Biosignal Interpretation* (pp. 219–222).

NASA - Mission Operations Directorate Space flight Training Management Office. (2009). *International Space Station Astronaut Candidate Training Catalog* (No. JSC-36543). Houston, Texas.

Panina, D., & Aiello, J. R. (2005). Acceptance of electronic monitoring and its consequences in different cultural contexts: A conceptual model. *Journal of International Management*, *11*(2), 269-292.

Park, S., & Catrambone, R. (2007). Social Facilitation Effects of Virtual Humans. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(6), 1054-1060.

Rasmussen, J., & Jensen, A. (1974). Mental procedures in real-life tasks: A case study of electronic trouble shooting. *Ergonomics*, *17*(3), 293–307.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE transactions on systems, man, and cybernetics*, *22*(6), 1296–1308.

Rowe, A. L., & Cooke, N. J. (1995). Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quarterly, 6*(3), 243-255.

Rubin, V. L., Stanton, J. M., & Liddy, E. D. (2004). Discerning emotions in texts. In *The AAAI Symposium on Exploring Attitude and Affect in Text (AAAI-EAAT)*.

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organizations*. Ablex Norwood, NJ.

Sexton, J. B., & Helmreich, R. L. (2003). Using language in the cockpit: Relationships with workload and performance. In R. Dietrich & T. von Meltzer (Eds.), *Communication in high risk environments* (Vol. 12, pp. 57–74). Helmut Buske Verlag.

Shriberg, E. (2005). Spontaneous speech: How people really talk and why engineers should care. In *Ninth European Conference on Speech Communication and Technology*.

Smith-Jentsch, K. A., Campbell, G. E., Milanovich, D. M., & Reynolds, A. M. (2001). Measuring Teamwork Mental Models to Support Training Needs Assessment, Development, and Evaluation: Two Empirical Studies. *Journal of Organizational Behavior*, *22*(2), 179-194.

Staggers, N., & Norcio, A. F. (1993). Mental models: concepts for human-computer interaction research. *International Journal of Man-machine studies*, *38*(4), 587–605.

Stanton, J. M., & Julian, A. L. (2002). The impact of electronic monitoring on quality and quantity of performance. *Computers in Human Behavior, 18*(1), 85-101.

Stanton, J. M., & Stam, K. R. (2006). *The Visible Employee: Using Workplace Monitoring and Surveillance to Protect Information Assets-Without Compromising Employee Privacy or Trust*. Medford, NJ: Information Today, Inc.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, Shared Mental Models, and Coordinated Performance: An Empirical Link Is Established. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41*(1), 61-71.

Strapparava, C., & Mihalcea, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 1556-1560). Fortaleza, Ceara, Brazil: ACM.

Vandenplas-Holper, H. C. (1996). Intra-individual and inter-individual cognitive conflict, related variables, and relations with cognitive development. *Swiss Journal of Psychology*, *55*, 161–175.

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, *48*(9), 1162–1181.

Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, *67*(10), 870-886.

Watson, A. M. (2008, March 30). *Electronic Monitoring Relevance and Justification: Implications for Procedural Justice and Satisfaction.* North Carolina State University.

Weitz, S. (1972). Attitude, voice, and behavior: A repressed affect model of interracial interaction. *Journal of Personality and Social Psychology*, *24*(1), 14-21. doi:10.1037/h0033383

Wells, D. L., Moorman, R. H., & Werner, J. M. (2007). The impact of the perceived purpose of electronic performance monitoring on an array of attitudinal variables. *Human Resource Development Quarterly*, *18*(1), 121.

Winitzky, N., Kauchak, D., & Kelly, M. (1994). Measuring teachers' structural knowledge.

Teaching and Teacher Education, *10*(2), 125–139.

Zhe, X., & Boucouvalas, A. C. (2002). Text-to-emotion engine for real time internet communication. In *Proceedings of International Symposium on Communication Systems, Networks and DSPs* (pp. 164–168).

Operational Assessment Recommendations: Current Potential and Advanced Research Directions for Virtual Worlds as Long Duration Space Flight Countermeasures

By
Dr. Jacquelyn Ford Morie
Dr. Gustav Verhulsdonck
Dr. Rita Lauria, J.D.

Wyle Laboratories- Final Deliverable

Contact:
Dr. Jacquelyn Ford Morie
email: morie@ict.usc.edu
phone: (310)-574 5708

Institute for Creative Technologies,
University of Southern California
12015 Waterfront Drive
Playa Vista, CA 90094-2536

Executive Summary

This report summarizes findings and presents recommendations based on a year-long research effort on the benefits of virtual worlds (VWs) for astronaut and ground crew training for long duration space flight. The components of this research included a literature review on existing applications of virtual world technology to space flight, an operational assessment on the most promising aspects of using virtual worlds for training for long duration space flight, interviews with a panel of six NASA experts on the use of virtual worlds, and participation in a three-day Behavioral Health Program Research Element workshop.

Based on the above components, this report summarizes the main findings of each of these components in order to describe current developments in virtual worlds and use of virtual technologies as applied to the context of long duration space flight. Furthermore, as a result of our research, we also assessed current training protocols at NASA and found exciting opportunities for augmenting existing practices for long duration space flight. Lastly, based on our assessment, we recommend promising directions of future research for NASA long duration space flight with an eye to implementation and cost-effectiveness.

Our findings are that virtual worlds are a promising technology for learning, preparing and supporting for long duration space flight. Briefly, a virtual world can be described as a virtual environment in which people may use avatars (virtual representations of themselves) to interact with each other and various elements of the environment. Similar to video games (which are usually more goal-oriented), virtual worlds may provide engaging ways of interactive and embodied learning that cannot easily be replicated in the real world. Given that current and future learners are part of a “digital generation” that is comfortable with learning, sharing and socializing through technology, we see a great benefit to adoption of virtual worlds for learning purposes (Prensky, 2005). As NASA increases cooperation with international space agencies, training is happening at a variety of different physical locales under increased time-constraining conditions. Although we do not propose to do away with face-to-face or physical (analog) training situations, we propose supplementing current training with the implementation of virtual worlds.

The benefit of virtual worlds is that they are cost-effective, malleable, interactive training environments that put astronauts into the various contexts they will be encountering while in long duration space flight. The benefits of this type of training versus traditional classroom training are as follows: 1) Portable and cost-effective and malleable 2) Interactive and Embodied 3) Safe 4) Non-collocated. Put differently, virtual worlds provide an easily configurable environment that provides astronauts with “hands-on” types of learning in an interactive manner. Furthermore, because they can replay and engage with learning material in a longer time-frame and on their own timeline, astronauts are given a safe environment to practice before entering actual analog environments. Coupled with video game-based types of learning approaches (where unlocking newer environments are based on achieving specific skills and goals), such learning environments may prove very effective for more comprehensive types of learning.

Furthermore, in international and dispersed training contexts, virtual worlds factor in more effectively by providing an online pedagogical and social backbone to the increasingly diverse international locales currently used in aeronautic training.

We see virtual worlds as a platform from which to support long duration flight missions in a variety of different manners. Below, we describe promising developments in artificial intelligence and online learning that may be combined with virtual worlds to support NASA countermeasure goals. Briefly, these countermeasures can be described as being comprised of a) Part-task training b) Academy, SFRM, Team-work and Ground Crew Efficiency Countermeasures c) Intelligent Agents/Tutoring Systems d) Resiliency and e) Delayed, asynchronous Decision-making systems.

One of the most promising aspects of virtual worlds is their malleable nature, allowing for an easily configurable, cheap and on-the-go online virtual training space. For part-task training, often involving routine or common operations, we see virtual worlds as incredibly promising in their ability to liberate existing analog (physical) simulations and providing a prolonged exposure to learning these operations through pre-analog, online digital training. In addition, by their nature, virtual worlds are easily reconfigurable and thus can be programmed to reflect the latest up-to-date versions of equipments. Likewise, we see great promise in using virtual worlds for physical training. For example, by coupling visual output to physical exercises, astronauts may train their vestibular systems to adjust to different types of micro-, low- and earth-gravity, which in turn would allow NASA to make a more well-informed decision in optimizing landing safety in various gravity situations and preventing unnecessary damage to flight equipment.

Another promising aspect of virtual worlds is their ability to create social networks and provide support in a variety of contexts. With NASA's cooperation with various international space agencies and increasingly globally dispersed training modules, greater support for individual, team and social interaction is required. Not only do astronauts need to "stay in touch" with their families, they also need to bond with other crewmembers if they are to function effectively in long duration space flight. As part of this, astronauts may utilize virtual worlds for various purposes, among which staying in touch with home, taking virtual vacations or using them to artificially create circadian rhythms and so feel connected to common rhythms of earth.

Furthermore, the use of artificial intelligence (AI) in conjunction with a virtual world (through for example *intelligent agents* or *intelligent tutoring systems*) may also provide learning of crucial communication, intercultural, decision-making and other types of tacit skills. Intelligent agents are virtual agents that use natural language processing (NLP) and so respond semantically to user input. Likewise, intelligent tutoring systems use such processing to create simulations wherein user skills are tested, evaluated and discussed post-session by way of a mentoring approach. In these simulations, the benefit of using a virtual world is that they create a safe environment to fine tune these skills and provide a specific means of testing and benchmarking various types of crucial skills. As long duration space flight will demand a greater emphasis on group cohesion, decision-making skills, intercultural/interpersonal and general communication skills, virtual worlds may provide key practice spaces for these types of tacit knowledge and experience. Furthermore, in creating more complex simulations using various intelligent components, astronauts and crew may encounter greater complexity and understanding of real-world contexts.

Another element that will be crucial for long duration space flight is training for resiliency – which can generally be defined as the ability to withstand hostile conditions and long lasting adversity by utilizing techniques to maintain strong and positive outlook on a group and individual level. Long duration space flight will be extremely taxing on astronauts and crew. Next to claustrophobia, social anxiety, feelings of isolation, loneliness and depression may jeopardize group or individual morale. We propose using virtual worlds to teach resiliency. For instance, next to allowing them to connect with “virtual buddies”, astronauts may connect (albeit asynchronously) with their family, ground crew or if deemed appropriate, the public in general through a virtual world. Knowing that they are not alone and keeping in touch will be paramount for astronauts, because it makes them aware when they need this type of relaxation and socialization, which will be crucial to mission success.

As communication with long duration space flights will be largely asynchronous (lasting up to at least a 20 minute time lag), it is crucial that flight crew and ground crew are trained in using delayed, asynchronous decision-making systems (DADs). Because there will not be an immediate ground crew response, astronauts will need to manage inquiries and troubleshoot largely on their own and decide on the appropriate course of action, simply because sometimes there is not enough time to wait for a response. In using a decision-making system, astronauts will need to learn how to diagnose and troubleshoot a particular problem, relegate a relevancy, and transmit the issue to NASA ground crew. After transmitting the issue, astronauts are asked to come up with their own solution by way of using the decision-making system, which may include intelligent agents, recommender and solution systems. The same time, once NASA ground crew receives their initial inquiry, they will need to start creating a solution as well, so that both flight and ground crew converge on the problem and a solution is found in the most efficient manner.

In order to prepare for long duration space flight, we see benefit in various efforts of research being funded in using virtual worlds. Not only are virtual worlds cost-effective, they are also malleable spaces that can effortlessly be used for online learning. Using intelligent agents and intelligent tutoring systems allows them to be used to train various hard-to-learn tacit skills by providing a safe environment that is highly configurable for a variety of skills and allows performance benchmarking. Research will need to be conducted in various areas, such as:

-Virtual world part-task training
-Virtual worlds for social networking and crew understanding
-Virtual world for online learning (intelligent agents and tutoring systems)
-Virtual world resiliency training (relaxation, charging up the well)
-Virtual world delayed, asynchronous decision-making (converging on a task, decision-making)

As virtual worlds start functioning as a transition space between human psyche and the outside physical world, they may also be suitable to transition us from the inner spaces of our mind to the outer reaches of our galaxy and beyond.

Table of Contents

1. Introduction

Cyberspace, that electronic space made up of myriad connections between people and information is *the* transition from space to outer space. It allows us to move from earthbound to the celestial terrains; proceed from the inner space of our minds to the outer space of the physical and beyond. In our networked, globally accessible society, the physical and the virtual continue to merge in ever more productive ways. Our mobile devices allow us to make use of an augmented reality wherein the features of the physical are explained in virtual terms through combinations of Global Positioning Systems (GPS), recommender systems, user-shared information, and immediately available information about our surroundings. We share personal information through social networks, play online video games, live in virtual worlds and increasingly, learn from networked environments and course management systems (CMS). We keep in touch with loved ones through instant messaging, email, voice and video chat. And this range of communication and information is no longer earthbound, as connectivity to the ISS and remote exploration vehicles has expanded our spheres of connectivity out into the near regions of our universe. Cyberspace technologies are rapidly evolving, and promise to expand our connections to, and exploration of, space in ways that are unimaginable today.

This report summarizes a year-long research effort on the potential of virtual worlds (VWs) for long duration space flights (for a glossary of terms, see the end of this document). Given the purview of this effort, this report makes recommendations on the most promising near-term areas of use for VWs within the particular context of pre-flight training, but also covers possibilities for during- and post-mission support. We see a tremendous potential here for these technologies to be used broadly, integrating them into part task training modules, enhancing team cohesion and intercultural understanding, and fostering psychological resilience across the spectrum of pre-flight training to post-mission support. VWs are also noted for their capabilities to be highly malleable and rapidly built, and therefore effective cost-saving technologies.

The data from this report was culled from the various stages of research, which were:

a) A literature review on available research concerning VWs and space flight

b) An evaluation of existing VW technologies

c) Qualitative data gathered by interviewing a panel of NASA operational experts

d) A three-day workshop conducted by the Behavioral Health & Performance Research Element

This report synthesizes our findings from the above to recommend the most promising directions for use of Virtual Worlds for countermeasures and future opportunities to enhance training for NASA space mission crews.

2. Terms and definitions

A Virtual Reality (VR) is a non-physical, created construct that provides an alternate environment for real humans to use and inhabit. While a VR can be an imaginary construct, for our purposes it comprises computer-mediated systems created especially to present such an alternative environment to a participant through one or more sensory channels. The term itself, Virtual Reality, was popularized in the late 1980s by an astute group of researchers and entrepreneurs working to ready the technology for popular usage. Sometimes terms from Science Fiction literature are used interchangeably with the idea of VR. Cyberspace is a term introduced by author William Gibson in his 1982 short story *Burning Chrome* (and used in his 1984 novel *Neuromancer*) to describe a connected "consensual hallucination" shared by real people inhabiting the virtual space inside the computer. Science fiction writer Neal Stephenson coined the word Metaverse in his 1992 book, *Snow Crash*, to describe a similar concept of realistic shared computer-generated spaces. Stephenson's term also connotes that these environments are a metaphorical representation of the universe and besides Gibson's psychological approach, thus also highlights the *spatial* and *metaphorical* dimensions of modeling space virtually.

The majority of early Virtual Reality applications were singular environments where one or at most a few people could share the computer world. They were typically single purpose, with some created for virtual travel such as a visit to the monuments of Egypt as they might have looked in their heyday, some promoted as exciting new game spaces and some even as artistic experiences. These Virtual Environments (VEs) were also explored for their more serious potential, including therapy to mitigate phobias such as fear of spiders, flying, heights, public speaking, as well as for training purposes, especially as promoted by NASA at its AMES Research Laboratories (Fisher et al., 1986). Using a VR environment, one could use virtual tools to affect a result at a distant place, where the actions in the virtual world controlled the motion of real tools at the other end.

Virtual Reality and VEs provide environments that afford a person a sense of immersion and presence in the artificial, computer-mediated space. Early versions of VR featured head-mounted displays that could insulate and sequester users within a digital space, and thus provide an authentic experience of being immersed and present in a different environment. Various other interface instruments, such as gloves and tracking devices were developed to traverse the virtual space or manipulate virtual objects. The tracking, manipulation and immersion constituted a distinct spatial experience, and for this reason, virtual environments are often considered spatial virtual reality. A spatial VR typically has limited or no connectivity to other VRs, is created for a single purpose, has a repeatable set of starting parameters (no persistence across usage) and minimal or no self-representation, and if it does, that representation is defined by the programs' authors and not the individual user.

Virtual Worlds, by contrast, are socially oriented Virtual Realities, built as persistent environments, or worlds, that may be inhabited, traversed, and manipulated by a person through their avatar (their virtual representation). They are characterized by multitudes of concurrent users, who may interact and communicate with each other while in the world. This persistence, as well as the multi-user aspect is a key differentiator of VEs from VWs. In providing extensive space and possibilities, virtual worlds can permit participants to model, to explore and to interact within a world that can be similar to our physical surroundings or which may present novel, imaginary, or fantastic environments that are not bound by physical reality. Unlike the physical universe, a VW suffers no such constraints, and therefore allows for a greater range of opportunities and experiences that may help prepare an individual to inhabit even further dimensions than the Earth and Cyberspace.

Video games share much with both Virtual Reality and Virtual Worlds. They provide common spaces for players to interact with each other in a virtual space while connected to a server. However, video games are based on a more goal-oriented environment in which achievements are emphasized.

For our purposes, both virtual worlds/environments and video games can be beneficial in the context of space travel countermeasures. Whether one learns, socializes or relaxes in a virtual world or in a goal-oriented game, we consider the experience more important than distinguishing whether the experience was defined as a virtual world or a game.

The popularity of video games as a popular pastime of many people is well known and publicized. The use of Virtual Worlds has also been steadily growing. According to GigOm Research, one in eight people in the United States report having used a Virtual World (Wagner Au, 2009). The use of VWs especially by the youngest demographic, kids from age 5 to 15, seems to be exponentially increasing (Marsh, 2010).

Indeed, between its game and virtual world use, the current generation can rightly be called a “digital generation” as they are comfortable in learning, living and communicating through computer games and computer-based environments (Prensky, 2005). The growth and prospects of the digital generation for NASA means that learning and communication through computers are familiar and expected ways of learning by current and future generations. Because of this, and for a number of other reasons, we see a tremendous opportunity for BHP to utilize VWs as digital learning environments in various contexts.

3. Methods

In our research towards the creation of this report, we utilized a variety of sources and methodologies to gather pertinent information: a literature review, an assessment of those findings, an interview component with a variety of NASA experts, and participation in a 3 day BHP Research Element Workshop. Based on these sources, we present our recommendations within the following areas:

- Virtual world and social networking potential for countermeasure training
- Gaps in pre-flight training that were found via research, interviews and the BHP workshop
- Potential future directions for research (basic and advanced)

Below, we summarize the findings from each method. For more in-depth information, throughout this report we will refer you to earlier work we conducted as part of this research effort.

3.1 Literature Review and Operational Assessment

For our literature review, we combed existing research and popular documents for anything related to virtual world technology in general, virtual reality (as used in the NASA space program), human computer-interaction, video games, social aspects of VWs, space travel and preparation for long duration mission (space and analog). This resulted in a good understanding of existing gaps in research pertaining to VWs and countermeasures, which are legion and will be discussed in the recommendations section as possible promising directions for new research for NASA. Other literature that was consulted focused on intercultural/ interpersonal and social human factors, theories of group management, crowd-sourcing, geographic information systems, telepresence, telemedicine and resilience. From this work, a number of different areas for countermeasure research were indicated that could be of potential benefit to NASA, with use of VWs for pre-mission training countermeasures deemed as the most promising.[1]

Therefore, for our operational assessment, we limited our research to the efficacy of VWs on pre-mission countermeasure training, leaving their use for just-in-time or during-mission support purposes to future reporting. We predominantly focused on pre-flight training and the different types of countermeasures that could be facilitated for psychological, physical and team performance. We divided the pre-flight training into individual, team and task-performance

[1] Full results of the literature search can be found in our first report: Morie, Verhulsdonck & Lauria 2010a

components. Based on our assessment, we described ways in which virtual worlds could facilitate several learning forms. Briefly, these include a) social and game-based digital training that provides active learning situations with emotionally resonant and embodied meaningful experiences for students; b) intelligent tutoring systems that provide review and reflection on learning situations and hone in on underdeveloped learning areas; c) Embodied Conversational Agents, or ECAs, that serve as near-humans for various purposes in the virtual world, and d) artificially intelligent systems that can analyze performance and create benchmarks that are connected to databases of reconfigurable scenarios.

We recognized additional factors that can influence an ASCAN's physical and mental health, such as individual well-being, team well-being, interpersonal relations, intercultural factors, and group morale, as well as the impact of micro-gravity on individual physical abilities (hemodynamics, bone-density, vestibular atrophy, "space brain" etc.). Countermeasure training must include methods of coping with such factors, among them a variety of psychological elements such as:

- General feelings of anxiety, social isolation, claustrophobia, depression (caused by being away from home and family, loss of circadian rhythm, etc.)
- Interpersonal and intercultural team issues (as a result of increased duration of flight and recent increased international cooperation between various space agencies)
- Group and team dynamics (leadership skills, team communication, task performance, team socializing)
- Individual health/psychological dynamics (assessing and recognizing one's mental and physical health and caring for the self when appropriate)

Based on the idea that VWs should help mitigate these factors, we concentrated our interviews and resultant operational assessment on finding current means of training that could help astronauts and ground crew deal with these issues in productive ways. Furthermore, we also looked at existing NASA training protocols (such as Space Flight Resource Management, part-task training, and various flight simulators) to locate potential gaps and means of improving astronaut and ground crew training for long duration space flight through the use of VWs. We identified a number of different gaps that we describe below.

Firstly, we identify VWs as extremely cost-effective simulations and digital training environments for astronauts. Compared to available physical analogs (e.g. physical simulations in a room), or high-priced Virtual Reality simulators, VWs are malleable virtual environments that are easily reconfigurable, accessible from any computer with an Internet connection, and provide multi-modal, *embodied* learning environments for ASCAN training. Current and future training environments can be easily converted to or created within a VW to address a wide range of situations. For example, a VW could simulate some of the salient effects of micro-gravity via virtual world experience and feedback and thus train the spatial and visual sense of an astronaut without requiring physical micro-gravity conditions.

Secondly, due to developments in social networking, we note that VWs are also well situated for creating a social network that can enhance team cohesion and allow for pre-mission bonding. For instance, astronauts could meet regularly in a VW and get to know the other people they will be working with during the mission. These meetings would include the participants taking on the form of an avatar (a personalized 3D representation of the self) within the VW. Research shows that people bond with this virtual representation and "act" through it, revealing many aspects of their social selves and developing a better understanding of the self (Yee & Bailenson, 2009). As future space crews are expected to comprise diverse international members, VWs and social networks are ideal means for astronauts and ground crew to become familiar with the social and cultural norms of their future colleagues, in addition to practicing anticipated shared tasks.

Thirdly, coupling VWs with artificial intelligence (AI) and Natural Language Processing (NLP), can make VWs highly effective online learning spaces. The use of Embodied Conversational Agents (ECAs) with AI will lead to the creation of more high fidelity and life-like social simulations. Further, Intelligent Tutoring Systems (ITS) are especially promising for online learning purposes, as they can provide student feedback and post-performance assessment. With use of VWs, NLP, ECAs, and ITS, all combined in a training simulation, NASA can not only impart high bandwidth knowledge, they can track the learning of both first year candidates (ASCANs) and astronauts and set benchmarks for future performance. For future purposes, when discussing the potential of VWs, in this report we refer to both groups under the inclusive umbrella term of "astronauts".

For example, astronauts can be presented with an off-nominal/nominal situation using a scenario that tests their knowledge, allows them to interact with ECAs via NLP, tracks their actions, provides just-in-time feedback via intelligent agents and game state information, and follows up with personalized post-session tutoring. Through engagement, personal feedback and the ability to "try it again" (thus learning by non-disastrous failures), Astronauts may prolong their exposure to learning material, and thereby create a more comprehensive cognitive model of various situation. In turn, NASA mentors and trainers can analyze the data from sessions and share their findings with the ASCAN.

Fourthly, we believe that VWs will function as social, entertaining and therapeutic environments wherein an individual may temporarily "escape" their immediate reality for rest, relaxation or recuperative purposes. These environments could provide virtual vacations, relaxation programs and techniques to help regulate sleep, alleviating some of the need for pharmacological agents. They may also include communication connections with friends and family, whereby both parties interact via avatars or "virtual humans" in an embodied manner. We think the latter aspect will be increasingly important in the light of long duration space flight, where astronauts need to develop ways of dealing with isolation, homesickness and cramped conditions.

Lastly, we see great promise for resiliency training in VWs. Resiliency training differs from traditional stress-training in that it explicitly develops an individual's ability to deal with adverse conditions in physical, psychological and social contexts through specific techniques. While currently used primarily in sports training for elite athletes, these techniques can be adapted to dealing with the adverse conditions of long duration space flight. Resiliency training focuses on instilling constructive, implicit methods that help individuals mitigate the effects of personal trauma through positive thinking, taking care of physical and psychological needs, and creating positive group interaction. We believe resilience will become increasingly more important in NASA training for long duration flight where conditions will be extreme and astronauts will be in isolation and cramped space facing hostile environments. Like athletes preparing mentally for a marathon, training in resilience techniques will enable them to create effective strategies to deal with hardships they are expected to encounter during their missions.

3.2 Interviews

Interviews were conducted between June 1, 2010 and June 8, 2010 with six individuals representing different areas of operational expertise within NASA (Morie, Lauria & Verhulsdonck, 2010c). The purpose of the interviews was to operationally assess the potential of using virtual worlds to augment current NASA training for astronaut candidates and flight controllers. The interviewees are briefly described below with their names withheld.

Interviewee 1 continues to support human space flight programs and ongoing research projects. He is actively involved in various human space flight programs, including commercial space flight. He does business development in terms of real mission support, for example, when new hardware is being developed, or when malfunctions with gear on orbit occur that require an early crew evaluation.

Interviewee 2 is a high functionary who works inside the Mission Operations Directorate. She is involved in diverse activities, including developing metrics for feedback, training and analyses, developing the training flows across areas like the Training Academy Course, common content for space station flight controllers, Space Flight Resource Management (SFRM) skills, and deals with the training program for astronauts and flight controllers for both the Shuttle and for the ISS.

Interviewee 3 is a psychologist with ground crew experience who works in the training standards integration group. He states this group basically does instructional design, curriculum development, evaluations, bench marking, and policy making for training in terms of the different divisions.

Interviewee 4 is a senior scientist and scientific advisor for Space Life Sciences and has extensive experience in academia and medicine.

Interviewee 5 is an ASCAN who background includes training as a medical doctor for emergency medicine and aerospace medicine. He has also worked at NASA as a flight surgeon and is currently involved with candidate training, which involves robotic arm training, spacewalk and EVA training, Russian language, flight training, flying in the T38, and space systems training.

Interviewee 6 is a senior scientist on the operational side of BHP and a representative from SFRM. She develops training for flight controllers and astronauts regarding the competencies in the SFRM area. Also, she is a participant in the international working group to design the SFRM competencies that they consider necessary for long duration missions and did the job analysis associated with those.

During the interviews, interviewees were asked open-ended question regarding the potential of VWs for training areas. Important aspects that interviewees brought up were practical considerations that exist during current and recent ASCAN training, such as the demanding travel and time commitments, increased cooperation with international space agencies, and various other aspects that long duration space flight will require. Interviewees also commented on their visions for using VWs in ASCAN training contexts for long duration space flight, noting especially the following areas: VWs used as “mini-sims”, and part-task trainers, training for leadership, training for boredom, VWs to help with loneliness, isolation, stress, relaxation, as well their use for on demand training and refresher courses. Key quotes from the various interviews are synthesized below in the results and discussion and integrated to provide a field perspective on the recommendations that follow.

4. Results and Discussion

As mentioned, the focus of this report is on discussing Virtual World countermeasures for pre-mission training. However, we do believe that many of the pre-mission training technologies could be carried over into just-in-time cross training countermeasures and support/adaptation countermeasures.

Our research points to a variety of promising areas where VWs could be used for countermeasure purposes to support ASCAN and astronaut task learning, social interaction, and group cohesion. The following elements seem particularly promising when combined with VW technology:

1) Part-Task Training Countermeasures
    a. VWs as portable, hands-on simulations
    b. VWs and Intelligent Tutoring Systems
    c. VWs and Embodied Conversational Agents
2) Academy, SFRM, Team-work, and Ground Crew Efficiency Countermeasures
3) Physical Countermeasures
4) Resiliency Countermeasures
5) Delayed, Asynchronous Decision-Making Countermeasures

We will describe each of these elements in further detail.

4.1 Part-Task Training Countermeasures

A common element of many professions is learning the "tricks of the trade". Even though many of us are prepared for particular situations in theory, important learning experiences are also conveyed through experience gained on the job. Oftentimes, this experience comes from exposure to real-life situations and provides a more coherent framework from which professionals can draw when making decisions. At times, such a framework is built upon a mix of tacit knowledge, muscle memory and oversight as a result of prolonged exposure to particular situations. An appealing aspect of VWs is that they can provide in-depth ways of providing embodied experiences that are much more engaging than many traditional and passive forms of gaining knowledge (such as lectures, books, and manuals) by giving people direct, hands-on experience in a learning domain.

Part-task analog training methods currently used by NASA to train astronauts include custom-built physical objects and settings that mimic instrument panels, vehicle sections and tool mock-ups. While we do not suggest that these methods be replaced, we certainly think that VW technology may offer a cost-effective, non-physical, widely accessible method for training both astronauts and ground crew, precisely because VWs do not require physical presence and can be utilized from every desktop. VW simulations can present important information about physical location and accessibility of instrument panel elements, allow for kinesthetic training, and provide both groups with feedback on consequences of taking particular choices, all without the need for reserving time to practice in a physical simulation room. In other words, VWs provide embodied learning environments that are cost-effective and portable, pre-training astronauts and ground crew in a more-in depth manner before heading into analog training situations.

A promising aspect of VWs is that they are suitable for digital, game-based learning. In many cases, this factor is a big selling point for trainees who are eager to get hands-on experience, but also want the training to be engaging. The difference between game-based VW learning and book learning lies in the *immediacy* by way of direct feedback that is experienced by people within the VW. Rather than reading about a situation, one is experiencing a situation in all its complexity and seeing the consequences of one's actions in a dynamic manner, which leaves greater sensorial and emotional impressions on individuals. For instance, interviewee 1, a former astronaut, noted their preference for interactive, rather than passive, forms of learning:

> The hands on things were always the most engaging, and I wager for all astronauts. People have intellectual curiosity, and they'll enjoy a lecture, but to have a hands-on simulation where you are actually participating as opposed to being just in receive mode, those are the things almost universally people prefer.

Through a VW, astronauts and ground crew may become aware of the consequences of their actions through various forms of game state feedback, direct feedback (for instance, through a mentor that comments on their choices), or feedback via artificial intelligence as the VW presents new challenges to overcome. Moreover, if a simulation utilizes narrative-based, emotionally resonant moments, a mental state of "being there" is formed that leaves astronauts and ground crew with important, deeply felt experiences which are retained more easily than the abstract scenarios offered by simple book-learning.

Another element of VWs that is positive for astronaut training is their ability to turn lost moments into learning moments by allowing students to explore learning material outside of classroom times in a playful manner on their own, if they wish. Pre-mission training requires NASA personnel to go through rigorous and time-consuming training schedules, and we think that VWs may offer astronauts and ground crew a more comprehensive means of understanding particular situations before entering real-time situations. For example, rather than constraining people to the analog, physical dimensions of a class or training room, the portable qualities of VWs allows astronauts to "play around" with a training room simulation before entering and so get a clearer idea of the different dimensions of a learning domain. A VW simulation can thus also serve to "ramp-up" astronaut and ground crew knowledge before they encounter the actual analog component. Astronauts and ground crew can thus maximize their classroom time, or more expensive training time, by engaging with learning matter at a deeper level in a safe environment. VWs can provide repetition of the material learned as well, giving astronauts and ground crew more opportunities to fully explore various aspects of the simulation on their own before, or after being tested in actual analog, physical training rooms. Perhaps as importantly, because VWs are easily configurable, they can easily be modified to reflect novel and challenging, or off-nominal situations, as needed.

Below, we will be discussing the benefits of using VWs as experiential learning systems in conjunction with intelligent tutoring systems, embodied conversational agents and embedded assessment methodologies.

4.2.1 Intelligent Tutoring Systems

One of the most effective ways to reinforce experiential learning is when a knowledgeable mentor is available for a post-mortem reflection, so the trainee can see through expert eyes what has gone well and what needs improvement. Though this type of learning has proven effectiveness, it is often difficult to have experts available after every hands-on lesson.

AI systems are advancing concurrently with the growth of VWs and their combination promises to be especially powerful. Intelligent tutoring systems (ITS) can now be utilized within real-time learning simulations to provide important feedback regarding student performance in the absence of an instructor, thus reinforcing and fine-tuning particular skills. Briefly, ITS can be described as (usually) a text-based artificial intelligence systems that provides learning moments by reviewing the performance of a participant in real time within a simulation, or by conducting post-performance reflection.[2] An ITS simulation can also be coupled to a high-fidelity Virtual World, adding a great deal of value and reinforcement. An example of ITS use in a VW might be in guiding a person through making a difficult ethical or personal decision that affects a group of people, by offering insights, suggestions or specific techniques that can be applied to the situation when, say, an incorrect choice has been made, or an unexpected turn has been taken.

Ideally, an ITS system presents a student with a contextual framework from which to operate (for instance, an understanding of particular issues and knowledge required for their choice), lets a student test out their knowledge in a practical situation, and then provides the student with feedback and reflection on their performance and the decisions they made. Natural Language Processing (NLP) is a key element of ITS, which uses either audio or text input, matching student output semantically with questions or suggestions directly related to the performance. The ability of ITS systems to create such responses is an important asset, as teaching skills are often difficult to gain without person-to-person, scenario-based exposure, including various communication skills such as conflict resolution. ITS can thus impart tacit knowledge skills that are difficult to learn on one’s own, but are critical to mission success.

When interviewed, interviewee 6, senior scientist for BHP Operations who has designed many simulations for training purposes, expressed her hope that ITS will be integrated with other learning systems in the future and that she was “hopeful” about developments in this field. Based

[2] For in-depth discussions of ITS see Morie, Verhulsdonck & Lauria 2010b

on our research in Morie, Verhulsdonck & Lauria (2010b), we see the area of ITS as a promising area of development for ASCAN training, because it provides experiential practice for hard-to-learn interpersonal "soft" skills that can help prevent problems between staff and create lasting impressions for understanding in situations where proper communication and decision-making is key.

4.2.2 Intelligent Agents

Another promising research area is the development and implementation of artificially intelligent agents that, unlike ITS, are embodied in some way. That is they are presented to the learner not simply as text, but as a character that imparts the knowledge via conversational means. Such agents are referred to as Embodied Conversational Agents, or ECAs (Cassell et al, 2001). They are extremely useful within socially-oriented VWs, in that they can be quite expressive, in their looks, emotions and movements. Such agents can interact with trainees in many ways and thus create a more complex simulation for training. Current video games represent just the tip of the iceberg when it comes to developing intelligent agents that can help in providing complex learning situations. For instance, subsequent installments of the game Half Life 2 feature a protagonist who is aided and addressed by a female intelligent agent who provides objectives, on-the-go tutorials, and generally offers assistance during intense situations. Similarly, ECAs could be very helpful in providing astronauts and ground crew with learning objectives, complex situations, and a practical application of their skills in basic and advanced simulations as part of a specialized and targeted social environment.

For instance, interviewee 2, a high ranking official who works inside the Mission Operations Directorate, expressed great hope in using intelligent agents to teach key skill sets that are currently not easily available to astronauts but that would be a great help:

> I could see situations that I could put somebody in and have them interact with an intelligent agent to get some learning objectives I wanted. [For example] I could put somebody in a situation and I could take my intelligent agents and give them the characteristics I wanted this person to deal with. Say I wanted to teach somebody conflict resolution and have them practice conflict resolution in a virtual world. I could give my intelligent agent certain characteristics. [For instance] they could be a difficult person, and I could get all of that into the intelligent agent. I could put folks into different situations because I could program my intelligent agents to have those particular behaviors that they may not get in the real world because they may not have that combination in any given real world situation.

Furthermore, the interviewee was convinced that these types of agents could be used both for basic ("fundamentals") and advanced training purposes. As most people are aware, real-time situations are often complex, messy and impacted by various elements such as group interactions, and VWs coupled with ECAs offer an excellent "testing ground" to a person who can thereby experience such situations in an emotionally resonant and pedagogically meaningful way.

Currently, NASA has several facilities to train flight controllers. Training takes place incrementally using analog, physical facilities. Part of the promise of VWs and intelligent agents would be their ability to make ASCAN flight training more portable and less time-constrained. Pre-training for analog situations could be facilitated, and more complex situations such as malfunctions could also be provided *prior* to ASCAN arrival at analog training facilities. For instance, a VW could provide an ASCAN with an off-nominal situation where various intelligent agents may, at times, interact with the person being trained, and provide differing accounts on what needs to be done. When asked what types of training could be facilitated by VWs, interviewee 2 further expressed their idea of how this type of training could benefit and augment current ASCAN training:

> We have several different facilities we use to train our flight controllers. And they start from having one or two people talking about a specific system. And they may be working with that system, and calling up their displays for that system, and getting to know them, and looking at their procedures. But it's just people in that system. And then it will grow, and they'll have say, 3 more systems, and a flight director. Now you have a bigger team working together, and they're engaging in nominal procedures, and malfunctions that happen, and their malfunction response. And then we get even bigger and get the whole team together who would be involved in supporting either a Shuttle mission or a space station increment. So at any of those points, especially the first two, I think we could put that into a virtual world and practice. I mean even if it is just one person with some intelligent agents at the other positions. They have an opportunity to practice their malfunction response. They can practice their procedures and going through their procedures, and their procedure familiarity. There [are] a lot of things that we could do in a virtual world, even at an advanced stage I think.

When asked what a good application would be for virtual world training, interviewee 2 answered that she favored "mini sims":

> Just putting them virtually in their work environment and allowing them to work through scenarios, either with 1) intelligent agents; or 2) with other flight controllers there with them so they are all in the virtual world. So they can just practice those things like I mentioned before: running their nominal tasks, running their procedure, malfunction recognition and response.

Even further, the same interviewee saw the potential of using VWs and intelligent agents (or other astronauts, ground crew) as a means of benchmarking crucial SFRM skills: "So I think for the ASCANs if we could develop some in station or in spacecraft scenarios that they can work or interact with other ASCANs or intelligent agents that we could put them in an environment that allows us to assess their SFRM skills in a real environment better." Through such "hands-on" situations, we think that many crucial skills can be developed if NASA were to teach SFRM skills by creating virtual worlds with intelligent components such as ECAs and intelligent tutoring systems.

In other words, there are great opportunities here for NASA to create benchmarks components within VWs for ASCAN performance on a number of different elements:

- General SFRM skills (Spatiotemporal skills, Nominal/Off-Nominal Task Skills, Diagnostic/Malfunction Skills)
- Communication "soft" skills (Problem and Conflict resolution, Interpersonal/Intercultural skills)
- Leadership and team skills (Task diagnosis, management, execution).
- Individual ability to deal with stress for varying duration (further discussed under "resilience")
- Technical tasks (Maintenance, Space walks, Equipment Repair/Troubleshooting)

Currently, some of these skills are taught by way of in-person board games or role-playing requiring people to be physically present. Likewise, instrument training is done in analog training facilities. This makes training for these types of skills reliant upon scheduling facilities and multiple astronauts to be present. While we certainly do not suggest that the traditional elements of training be replaced, we do think that using VWs with ECAs and ITS can be a valuable supplement to such training, allowing, as it does, additional personal practice that does not require scheduling of physical assets, or coordination of trainees, or role-players.

Next to their ability to be portable, active and "hands-on" training, VWs can also benefit astronauts and ground crew by providing a safe, personalized learning environment that they can access at any time. A real potential for improving one's skills are that VWs provide a safe environment in which to test those skills without fear of social ostracizing or repercussions due to failure, or risky outside approaches to challenging situations. Because they can practice at will, and perhaps take bigger risks, students may gain a deeper understanding of complex skills that can support the challenges of long duration space flight.

As we discovered during our interviews, current SFRM training is tightly scheduled and uses learning flows, but there are few, if any, 24/7 training facilities that function as personalized learning environments for astronauts. Current developments in online learning show that course management systems can be effective for long distance learning and minimizing the requirement for classroom presence. Like online learning, one of the obvious benefits of VWs is their accessibility for facilitating personalized types and on-time online learning. When asked if VWs were to be incorporated as supplemental training and whether they could accelerate training, interviewee 3, a psychologist who works in the training standards integration group, affirmed that it could have potential benefits:

> Now, I agree it would increase their tacit knowledge. Hey, you're part of the experience of it. Now you have experience as opposed to none. I don't know how fast that would be in terms of speedup their knowledge. But in terms of convenience and improving their tacit knowledge, that would be improved because you would be able to access this simulation based upon the individual student's timeline as opposed to making sure it follows a particular flow with an instructor.

In conclusion to this section, we see not only a benefit in providing VWs as part-task training simulations, but as a way of perfecting core SFRM and other skills by giving astronauts opportunities to work on these skills individually on their own or through scheduled time with the in-world VW learning. Furthermore, working on personal, social skills may be less awkward for astronauts in VWs and may bolster their ability to perform in face-to-face contexts by giving them the opportunity to train for these crucial skills in the safety of a virtual world with intelligent agents. This type of learning may be more motivating for some people to learn highly valuable skills that they may personally feel they lack without fear of social penalty. Coupled with intelligent tutoring, these systems can help astronauts gain greater insight and help them hone in on and rehearse various areas in which they need improvement.

5. Academy, SFRM, Team-work & Ground Control Efficiency Countermeasures

5.1 Virtual World Social Interaction and Intercultural Issues

As part of long duration flight mission preparation, we also foresee a great need for international, long duration crews needing to "stay in touch" with colleagues and so develop opportunities to interact with each other socially and culturally pre-mission. People that have known each other for a long time are able to communicate with greater understanding due to shared history, affinity, and social bonding. We see this as crucial for long duration space flight crews, as they must rely greatly upon each other throughout the mission. VWs can facilitate social interactions by having astronauts and ground crew create avatars that resemble themselves ("veritars") that provide optimal remote (non-co-located) communication with fellow personnel. The obvious benefit of using the social network capacities of VWs (such as avatar co-presence, chat, instant-messaging) would be in creating greater group cohesion, sharing learning experiences, and getting to know other crew members socially and culturally before they embark on their mission.

Interaction between different cultures requires overcoming not only language barriers, but also social and personal cultural histories. According to Hall (1980, 1982), a difference exists between cultures that rely highly on historical customs and social cues to establish context (high-context) versus those cultures that use verbal expression to establish context (low-context). Roughly, Japanese culture can be identified as high-context in that they place high importance on social customs and gestures during communication, whereas cultures such as the United States and the Netherlands can be characterized as low to middle-context in valorizing verbal expression over custom to help establish context. To those not used to it, high-context culture may seem to favor ambiguous, distant and indirect communication, whereas low-context culture is more accessible but highly reliant upon verbal communication to clarify. As a result, different cultures communicate differently, which is crucial for ASCAN training. Likewise, differences exist between various cultures in people's level of comfort with ambiguity and uncertainty, hierarchy, social position and community (Hofstede, 1980, 2001). These intercultural factors will most definitely affect group communication in future international crews and require cultural sensitivity and social acclimatizing that can be rehearsed and learned through VW training.

In conclusion, as tight training schedules in multiple locations and international diversity of future crews exist and may likely increase, we foresee the importance of VWs for providing a much needed infrastructure and framework for social interaction and enhancing group communication and understanding.

6. Physical Countermeasures

During our research, we also focused on the physical training that would be necessary for long duration flight. Astronauts will need to be trained for particularly strenuous conditions in which their physical state is severely tested. Much of this countermeasure training requires understanding of the medical consequences of micro-gravity on the human body. Briefly, the most serious of these conditions are loss of bone density and muscle mass, and vestibular and hemodynamic atrophying ("space brain") affecting acuity, cognitive functions, balance and spatial orientation. For this purpose, we recommend that NASA astronaut training involve benchmarking of balance performances, which can be accomplished through coupling VWs to haptic feedback (resistance or touch feedback). This can provide ongoing measurements and insights to medical personnel monitoring the crew members' physical abilities, and help indicate their readiness to perform certain key tasks for long duration space flight.

Ascertaining current physical performance within new gravities and environments can be simulated via virtual worlds and can provide valuable metrics. As interviewee 4, a senior scientist within Space Life Sciences remarks:

> [People] evolved for 24 hours in a day, in 9.8 meters/second^2 - the acceleration for gravity. And virtually all life we know and understand on this planet evolved in that. So we, as best as we can understand, have no pre-formed design to adapt to a lower gravity. We have some design for hyper G, but not for lower G.

An important part of long duration mission success thus rests upon developing models that can simulate these conditions. Part of the process of "adapting" to such an environment may come from modeling physical conditions in a VW as a transition space. Presenting a visual analogue to tasks that astronauts are required to perform in low-gravity or micro-gravity environments, VWs may provide a good modeling space to train astronauts to adapt to such conditions through, for instance, a high-fidelity head-mount display or on a computer screen. Such methods would especially be useful when combined with physical tasks (i.e. moving an object), because they help people to adapt and transition to these conditions. Because VWs are malleable, these conditions can also be quickly changed to prepare the shift from micro-gravity to low-gravity by, for instance, providing a greater sensorial experience to astronauts of object weight through haptic feedback when doing manual tasks. We believe such VW training presents an excellent manner to benchmark and monitor physical performances that are crucial to long duration mission success.

7. In-flight Resiliency Countermeasures

In addition to physical health, long duration missions also require great *resiliency* on behalf of crew members while in-flight. For this reason, pre-flight countermeasures should emphasize successful techniques that may help create *individual* and *group* resiliency. Akin to an athlete training for a long marathon, the concept of resiliency is based on positive thoughts, effectively coping with stress, psychological self-protection, individual fitness, and generally making sure that one is prepared for adversity. Resiliency training is used by the military to prepare troops for coping with the extreme and trying conditions of war so as to avoid trauma by giving individuals skills to be emotionally, physically and psychologically resilient to long-lasting adversity and duress.

We believe that long duration space flight also requires astronauts to train for resiliency. Individuals selected for long duration missions will need a complex skills-set that includes vast amounts of intellectual, individual and physical reserves, but also strong coping skills to deal with harsh conditions involving claustrophobia, social anxiety, and the high-stress of adverse environmental conditions. As explained by interviewee 1, a NASA employee working for Wyle in human space flight programs and ongoing research projects, currently NASA trains for high-stress decisions, but does not specifically focus on individual or team resiliency:

> I don't think there was any training that specifically catered to that [resiliency]. We do get a lot of training in stressful environments. Granted we don't do Shuttle mission simulations anymore, or at least won't once the Shuttle is gone, but that was very high stress. Alarms flashing and things breaking around you, and you had to recover from them in real time. Also in the training pool, they would simulate things breaking, and we would have to respond in real time based on our knowledge and skills.

In other words, while NASA simulates stressful situations and off-nominal/nominal malfunctions, a gap exists in astronaut training for resiliency that prepare flight and ground crew to be physically, intellectually and psychologically resilient and ready with techniques that can help them adjust quickly to hostile environments, troubling social situations and conditions that test their mental and physical reserves.

VWs can be designed to mimic harsh conditions and physical impacts visually, and can also provide means of tracking information about a person's mental and physical responses during the experience of such situations. Having this type of data can help determine a) information about the type of person and their natural tendencies; b) track progress and development of astronauts by way of ongoing data analysis; c) create benchmarks for individual resiliency and d) develop successful strategies and learning methods for astronauts to be more resilient over time. When asked, interviewee 4 mentioned how VWs could be used to train for resiliency:

> [As] I said you go through your benchmark series with a number of people. And subsequently put them on various types of strain. The sleep shift was one of the ones I told about. You could put them in a 6 degree down headrest setting, which gives some of analogue fluid shifts you have occurring in space. Then test their performances, and see how they do. And performances that are dexterous, not so dexterous, some that are solely cerebral, I think you would begin to find out where that resilience factor is and you might actually contribute more to understanding what you need to do for keeping people reasonable healthy and capable for very arduous missions.

Benchmarking is an important asset to NASA in establishing ways of ensuring mission success. By creating benchmarks, potential errors by personnel are minimized because benchmarks ensure the person who is performing the task has been trained and optimized to do so. As interviewee 4 mentions, benchmarking resiliency will play a large role if long duration missions are to be successful:

> The resilience side of this takes in a part that I have a personal interest in, no expertise but a personal interest. That is looking at performance. We can measure performance in a number of ways throughout training, and you do this. But the realities are when you fly a spacecraft like the Shuttle and you are up there for extended periods of like 17 or 18 days, which is a longer mission, and the Commander has to come back and land this spacecraft and actually operate that craft within minutes of reentering 1G. His vestibular system is not fully readapted yet. We are very fortunate because the landings have all been acceptable. So we haven't had anybody out of nominal range.

In other words, long duration missions will require a re-training and adaptation to gravity that can be done, in part, before the return to Earth through VW training. During the mission, flight crew could utilize spatial tasks to test their vestibular systems and so prepare for landing. By using VWs and coiled spring objects in space to train for these tasks, flight crew could re-adjust to gravity by helping them re-orientate to increased gravity after spending a long-time in micro-gravity. Another benefit to using VWs is that astronauts could receive a personalized training schedule that would provide them with direct biometric and visual feedback on their performance and so could help them improve their condition. Likewise, this data can be shared with ground crew to determine which person has adapted their vestibular system for optimal landing success.

7.1 Individual Resiliency Countermeasures

Next to benchmarking, VWs could be crucial for individual relaxation and entertainment purposes and provide another means of resiliency. Given the time lag, pre-recorded VW sessions could, for example, have mindfulness therapy to reduce stress. Current research at the Institute for Creative Technologies at the University of Southern California focuses on using VWs to treat soldiers with post-traumatic stress disorder, and more strides need to be made in finding out the potential of VWs in helping humans relax in otherwise stressful environments.

In order to further facilitate resiliency, our research indicates that other directions for VWs exist that potentially could be of benefit to flight crew. As a result of long duration space missions, flight crew could experience social anxiety, claustrophobia, and depression or simply miss being "back home". NASA understands the psychological impact of claustrophobic spaces and should look to the ability of VWs to mitigate such feelings. Briefly, examples of this type of research could be based on socially isolating crew and utilizing VWs to help them relieve boredom, anxiety and social phobias as a result of the isolation. A good example of the effects of an enclosed environment is the 2010 mining disaster in Chile, which left a group of miners in a small, enclosed space suffering from depression and anxiety. As of this writing, NASA was sending psychologists and other trained personnel to help out in the effort to keep the miners in good spirits until their retrieval from above ground. Being able to connect with the world (albeit virtually) will be an important aspect for long duration space flight.

Long duration space flight will not only create anxiety-inducing conditions, but also require adjusting to the loss of circadian rhythm, as a result of being in space. As any person who has experienced jet lag knows, the change in time severely affects functioning, and without a sense of time or gravity, sleep loss and no regular daytime rhythm creates less than optimal conditions for astronauts. Though we know that NASA research covers a wide range of ambient and mood-setting lighting to create an artificial day and night rhythm for astronauts, we see benefits to using VWs as means of complementing these efforts. For instance, while waking up, the astronaut could be seeing a VW in which the lighting creates a mood of early morning. For this purpose, we also further propose that VWs could provide real-world or imaginary spaces to explore if crew feel the need to escape for a moment, have intelligent agents that function as emotionally supporting "buddies" to flight crew, or flight crew could be provided with "virtual vacations" in virtual worlds. On the potential of virtual worlds for NASA, interviewee 4 particularly remarked the last element as promising:

> I can see how they [virtual worlds] would be useful in a number of settings, particularly for virtual vacations. You have people who are living in a can for 30 months and have no place to go so to speak. There is not a one of us who does not seek some type of release at some time or another to get away from everybody, everything. To go on a vacation, get away from everything, go diving, do whatever you want to do. I think these kinds of

> technologies will enable that to be possible. And people will get a lot of reinforcement from being able to depart, get diverted away from the missions, the tasks, all that stuff.

The ability to be resilient also depends on retreating at times to regain strength, new ideas, and energy. The importance of providing VWs to crew would be of tremendous benefit for relaxation, rejuvenating downtime, as well as a help to focus on non-work ideas and to enhance a sense of circadian rhythm. As in regular professions, such individual relaxation is of tremendous importance in order for people to refocus their efforts. In addition, VWs can provide social spaces for astronaut families to communicate in an asynchronous fashion with their loved ones in space.

8. Delayed, Asynchronous Decision-Making Countermeasures

Another element that will have a pronounced effect on long duration flights is delayed responses from ground crew and home. Not only does this create a psychological barrier between ground staff and the flight crew, it also creates a need for support systems that introduce more autonomy to those on the off-world mission. This includes dealing with functional failures without immediate intervention by ground crew, as well as autonomy in setting schedules according to the needs of the crew in space. Due to the great distance, communication lags are expected to last up to at least 20 minutes or more each way. As a result, any communication must be optimized and task-focused. According to interviewee 4, long duration flight missions require great mental reserves, including "psychophysical, decision-making processes, perception processes."

The consequences of long duration space flight communication delays place a greater importance on the *autonomy* of the flight crew's decision-making processes to resolve issues on their own without immediate assistance from ground crew. For this purpose, we also propose astronauts and ground crew interact with a VW system for decision-making that simulates this delay and tests their ability to make successful decisions. This can help both flight and ground crew to train for such situations, while also giving NASA researchers the ability to benchmark and optimize personnel performance related to delayed responses. We propose a VW-based delayed, asynchronous decision-making system (DADS).

We envision this delayed, asynchronous decision-making system comprising these elements, all involving a time lag for responses with ground crews and experts:

- A VW simulation that presents specific (known or expected) situations, as well as unexpected and particularly difficult situations that may cause the mission to fail.
- A decision tree system coupled to a database that allows astronauts and flight crew to diagnose and triangulate various types of problems. The paths through the decision tree can be designed to help in autonomous decision-making as well as to determine which ground crew subject matter expert technicians may be optimal for a particular problem

- The inclusion of intelligent agents to guide the diagnostic process, and suggest possibilities that might not occur to the on-board crew members.
- An exhaustive database composed of reconfigurable narrative elements coupled with the decision tree system dealing with as many as possible exigent malfunctions and a range of solutions to them
- A recommender system that can present possible solutions or suggest solution elements based on input to search algorithms

Current NASA training utilizes analog controls and crew feedback to practice for missions. For long duration purposes, we think it will be extremely beneficial to develop an expert decision-making system to diagnose a problem, to provide possible directions for decisions for astronauts, and to provide time-lagged input by ground crew. Again, the benefit here would be to benchmark quality of decisions and a means of creating well-informed decisions for flight and ground crew. Furthermore, if NASA were to couple the space craft working equipment to remote-sensing (such as Radio Frequency Identification (RFID) that would provide information on equipment components), ground crew would be able to pick up any equipment failure independently of flight crew, providing ground crew with a means to troubleshoot the issue while providing the flight crew with important information on how to resolve the issue.

Ideally, astronauts would be able to ask a query to ground crew with the sense of urgency, use the decision-making system while waiting for ground crew response, and determine what they need to do and be up to par with ground crew. If the decision-making system also streamed flight crew data from the DADS system to ground crew in a data stream, NASA would have an opportunity to see flight crew decisions, and at the same time, work on the initial query as well (if, for instance, the remote-sensing system had already notified them of the issue prior to flight crew's notice). Ideally, flight and ground crew data would be transmitted to each other asynchronously to be used as the crew sees fit. A typical protocol to a query could be described as follows:

1) Flight Crew
   Problem -> Initial Query with possible resolution and urgency level -> DADS

   Initial Query is sent to NASA with accurate problem description, and if possible, initial proposed resolution to treat problem (malfunction, repair) and urgency level. Flight crew then starts assessing potential solutions using DADS -> uses Decision-making System to refine Query and find possible resolution of problem.

   Query found in database and matches problem -> Solution found and if possible, await ground crew confirmation. If not possible, flight crew decision can be made with knowledge of best possible solution presented by the DADS system.

2) NASA Ground crew
   Based on Initial Query -> Confirm/Disconfirm or Wait for Further Instructions

   Based on initial problem description and proposed resolution, ground crew provides initial answer to proposed resolution (confirm/disconfirm) or hold for more information.

3) Wait for Further Instructions
   Ground crew sends in-depth instructions to crew on how to resolve problem. After Initial Query, NASA receives stream of flight crew decisions and gets a better idea of the problem domain and the crew's thinking.

   NASA crew, based on Initial Query and input information from DADS, formulates answer. If the problem requires in-depth knowledge, NASA can diagnose problem domain using input from ground-based experts in addition to the flight crew solutions based on their own knowledge and use of the DADS.

4) Flight crew combines own decision with that of ground crew for best possible problem resolution, if time permits.

Instead of overtly relying on ground-crew reliance, flight crews would be asked to take on flight problems and come to a resolution with the help of the delayed asynchronous decision-making system. In this situation, the decision-making system would help the flight crew and ground crew to resolve a problem and to do so by optimizing their time during the time lag. Rather than waiting for a problem to be solved by ground crew, flight crew would be able to figure out potential solutions to the problem themselves while waiting for a response. Coupled with the input of the NASA ground crew, the flight crew could then make an optimal decision. If the delay prevents this, then the DADS may offer the best immediate choice. In any case, we recommend extensive practice with delayed communication before flight, and optimized decision-making training through some type of DADS protocol when time issues preclude waiting for ground crew input.

9. Recommendations

In this section, we briefly make recommendations on the above results and discussion. Our focus in this section is to provide a roadmap for promising research areas for NASA. Rather than in-depth descriptions, the following should be seen as general directions of research that would be desirable for VW use in preparation for long duration space flight. We see great potential for research in the following areas, which could enhance NASA's current training procedures. Though we see all of these areas as important, we indicate the difficulty and depth of future research by mentioning whether this research is easily to implement or hard (in which case it is mentioned as Advanced research).

9.1 Research for creating Virtual Worlds as Online Part-Task Training Countermeasures

Virtual Worlds can mimic various flight, kinesthetic and operational procedures. NASA Part-Task Training can be conducted online in a virtual world before or after analog procedures. Potential technologies of interest for creating part-task training are portable, remotely accessible VW simulations using game-based approaches. Training scenarios should be narratively engaging, emotionally resonant and motivating, and function as interactive digital environments for astronauts and ground crew that can help track their progress and performance. Since Part-Task Training focuses on diverse elements, virtual world simulations need to focus on specific but diverse learning objectives for long duration space flight, among which may include:

- Standard Procedural operations
  - Flight preparation/ take-off / monitoring
  - Flight crew- Ground Crew interactions
  - Space Flight Resource Management

- Kinesthetic and vestibular performance in simulated micro- and low-gravity environments
  - Flight Landing/Spatial Orientation and Perception using micro, low and terrestrial gravity as environmental variables
  - In-flight object handling using coiled springs and visual VW feedback
  - Exercises facilitating cerebral adaptation and vestibular normalcy switching between micro-, low- and terrestrial gravity
  - Longitudinal sleep angle studies coupled with VW to investigate vestibular and hemodynamic consequence of performing spatial tasks that simulate low/micro/terrestrial gravity environments
- Part Task Training with team members to practice tasks that require coordination amongst crew members. This can be done even when team members are

geographically distant from one another, enhancing the operational effectiveness of the crew when they actually perform team tasks

- Malfunction responses to common nominal and off-nominal situations
  - Analytic ability/Troubleshooting
  - Decision-making
  - Communication
  - Task performance

9.2 Advanced Research for creating Intelligent Tutoring Systems and Intelligent Embodied Conversational Agents in Virtual Worlds for team-based countermeasures

Virtual worlds can incorporate artificial intelligence through intelligent tutoring systems and embodied interactive intelligent agents. Through use of natural language processing (NLP), such artificially intelligent elements can facilitate advanced simulations that go beyond procedural operations. Complex social situations where multiple elements of knowledge may be tested in a practical manner can be supported. Scenarios can include a cognitive framework and a simulation that tests ASCAN performance on different variables of "soft" skills:

- Communication skills
- Intercultural skills (High/Low-Context, Uncertainty Avoidance/ Ambiguity Comfort, Hierarchy, Social Position and Community)
- Analytic/Troubleshooting skills in complex and dynamic situations
- Performance under stress skills
- Decision-making skills
- Interpersonal skills (e.g. using Myers-Briggs Type Indicator test)
- Group Leadership/Team skills
- Task performance

As part of this work, simple simulations could be created to test for any of these soft skills. Advanced simulations could be developed to create scenarios for nominal and off-nominal tasks that support learning directed towards more complicated scenarios. These simulations would be comparable to commercial video games and include various branching scenarios, decision trees and AI that responds to ASCAN actions and responses. Complex, randomized algorithms can describe finite state operations and randomized states to create virtual world analogues to various real-world long duration flight for social and task-oriented situations.

9.3 Research in Virtual Worlds as Online, embodied social networks for astronaut and ground crew (Academy, SFRM, Team-work, In-flight Efficiency Countermeasures)

More focus on social connections and training through VWs may also create opportunities for group cohesion. Because long duration space flight places greater stress on flight crew, more emphasis will need to be put into team compatibility and cohesion. VWs may create team cohesion by proving astronauts and ground crew a means to meet other crew members virtually pre-flight, and so form social histories, deeper understanding of each other's personalities and communication styles, as well as knowledge of each individual's goals and objectives for the mission.

- Create avatars that resemble the physical person ("veritars")
- Greater emphasis on online astronaut and ground crew training through VWs
- Encourage sharing of learning experiences as astronauts and ground and socializing through a VW and so create a community of long duration space flight crew

However, when one astronaut was queried during the BHP Research Element Workshop as to the perceived benefits of this social enhancement to their training, they replied that astronauts would NOT do this if it were left to them to do it on their own time. The reason given was that every spare moment they have, they spend with their families. Therefore, for this to have the greatest benefit, virtual worlds must be emphasized as a key element within long duration space flight training flows. In turn, this strategic use of virtual worlds would create a social backbone to unite physically remote training colleagues that are living and travelling all over the world, and that must recognize and adapt to different cultural norms.

9.4 Advanced Research in using Virtual Worlds for Physical Countermeasures

Next to basic research in spatial orientation, advanced research for VWs for physical countermeasures could be utilized to study the effects of long duration conditions (triggering social anxiety, feelings of isolation, loss of circadian rhythm and others) by coupling VWs to analog, physical training exercises in confined spaces to achieve various goals. Potential goals of this type of research would include testing physical feedback given by a virtual world and providing feedback to astronauts on their performance. Part of these exercises would need to use coiled springs, or other methods, to provide friction for physical exercises that are measured and displayed in the virtual environment. Various research directions can be distinguished based on this idea, such as:

- Establish benchmarks for optimal physical, psychophysical and psychological performance in different kinds of gravity
- Establish protocols for providing biofeedback to trainees on aspects such as eye-movement, and physical and hemodynamic activities during vestibular exercises and create exercises tied to established benchmarks

- Artificial circadian rhythms that create an artificial day/night rhythms for astronauts to latch on to while on long duration flights, especially after "slam-shifts" interfere with normal sleeping patterns
- Comparative studies on difference in VW exposure versus regular, physical training protocols on bone density and muscle mass, acuity, cognitive functions, balance and spatial orientation (2x2 design, with control and experimental group using alternately, regular training protocol versus VW biofeedback exercises)
- Create advanced scenarios to test and create spatial and visual abilities using coiled springs and VW interface with audio-visual feedback

An important aspect for this would be to integrate and test how VWs may augment astronaut crew training by providing feedback to astronauts on their condition and performance during (immediate) and across (longitudinal) VW exercise regimes. This training would ideally be done in confined, isolated spaces such as those currently conducted by NASA in Alaska or off the coast of Florida underwater so that astronauts can prepare for optimizing their exercise routines while on long duration space flight missions.

9.5 Advanced Research in using VWs for Resiliency Countermeasures

Another promising direction for VW training is the addition of resiliency training to existing astronaut and ground crew training modules. In addition to standard training methods for nominal/off-nominal or stressful situations, creating resiliency techniques and learning methods using VWs can help flight and ground crew maintain a positive mental outlook and physical strength by providing life-long techniques that mitigate the formation of stress and trauma during missions. Promising areas for resiliency can be found in the following research:

- Psychological effects of keeping in touch with earth by communicating with family and ground crew through VW interface in the form of "embodied" asynchronous voice conversations with hugs/body language
- Virtual space "buddies" in the form of a VW intelligent agent that supports each astronaut and may help them discuss personal issues without fear of reproach by human team members using techniques and scenarios for "virtual" counseling
- Virtual vacations that give astronaut the opportunity to relax and escape from work by exploring VWs (could be coupled with entertainment, such as gaming, socializing with intelligent agents, or asynchronous communication with public), as well as with simulations of varied weather patterns.
- Mindfulness Stress Reduction techniques (such as meditation, yoga) delivered via a VW to help de-stress astronauts and maintain resiliency
- Virtual Reality Exposure Therapy (VRET) to help astronauts overcome or cope with adverse, chaotic or stressful situations they may encounter during a mission. This is a graduated re-exposure to the triggers that may have induced the initial stress reaction, and

is a standard, beneficial treatment for trauma survivors. Some form of familiarization with this technique should be included in pre-mission training.

As we found during our research, resiliency holds huge promise for mitigating the deleterious effects of long duration space flight, by unequivocal preparation for and understanding of how such stressful or compromising situations can be positively approached from the outset. Astronauts going on long duration space flights should be expected to be resistant against long exposure of the unfriendly and human adverse conditions of outer space as a matter of survival.

Due to the harsh conditions when travelling such long distances, it is difficult where to draw the line with practicing for resiliency. It is important to prepare astronauts in the best manner possible, but emphasis should be placed on group and individual survival skills in these environments. Ethical treatment of participants in this type of research is paramount, and standards will need to be put into place to ensure that participants are not harmed or unnecessarily traumatized while training for resiliency. Safety measures to prevent astronauts from getting hurt unnecessarily will need to be carefully stipulated and any research in this area must pass a medical and psychological ethics review board.

We can emphasize the importance of using VWs for positive ways of relaxing, de-stressing, socializing and vacationing, as well as for resiliency purposes, that pose little or no risk to astronaut trainees. Other protocols will need to be developed to help with the countermeasures that deal with the impact of long duration space flight on the physical body as discussed above, but these too will be important countermeasures for improving health during long duration space flights.

9.6 Advanced Research in Delayed, Asynchronous Decision-Making Countermeasures through a Virtual World

Advanced Research should focus on creating new methods for ground crew and flight crew to communicate in a VW simulation that factor in time lag of at 24 minutes or longer. Due to this factor, greater autonomy is expected of flight crew in determining solutions to problems after initial take-off. The delay requires simulating communication and decision-making in which dynamic problems need to be resolved in short order. Communication will require express operational commands to indicate whether a problem is nominal or off-nominal, the urgency, and proposed solution as given by flight crew. Depending on urgency, flight crew will need to assess whether they can wait on ground crew permission or if they can proceed in resolving the issue. They should be able to start problem solving using a delayed, asynchronous decision-making system (DADS) before the initial inquiry will even reach ground crew. Likewise, time will need to be optimized for ground crew to converge and reach a decision in assessing flight crew inquiry.

- Create/Implement remote-sensing communication protocol for long duration flight equipment (RFID to wireless to ground crew message of equipment failure)

- Create communication protocols for transmitting initial inquiry and proposed solution
- Create delayed, asynchronous decision-making system database that can be accessed by flight and ground crew and used for training and long duration mission purposes
- Create simulation with intelligent agents, and a recommender system for problems with an accessible database using optimization algorithms for astronauts to practice in nominal and off-nominal situations
- Utilize flight crew autonomy and analytic skills first and if necessary and possible ground crew input to decide and converge on a solution.
- Receiving and streaming solutions may help NASA utilize time productively and gather subject-matter expert technicians and solutions quickly by working concurrently on the problem domain together with flight crew
- Flight crew will need to be able to assess own solution and compare with NASA ground crew if time permits; else, proceed with autonomy and be assured of quality of their own decision by database recommendation

A VW may help astronauts practice these skills, especially if visual information can be tied to technical information. For instance, a powerful way of locating a malfunction in a flight cabin might be tying a virtual world model of the cabin to a database containing technical information about that particular area. If the area of the problem is unknown, a text database that functions to identify the problem and localize it in a particular area may also be useful. Protocols for this type of communication would need to be established and be based on the premise that flight crew solves most pressing problems on their own with minimal assistance of ground crew while less critical problems may require waiting for ground crew confirmation.

Conclusion

The efforts of this year-long research project focused on looking at ways astronaut and ground crew training for long duration missions could be enhanced through the use of virtual worlds. VWs provide various means of mimicking and simulating environments, afford social connectivity via avatar use and interactions, and can be very efficient and cost effective training modalities. Depending on future developments in wearable computing, flexible interfaces and processing speed, we may soon see complex augmented reality systems that form a virtual grid over the physical world. Current mobile technology is already at the forefront of such augmented reality (for instance, apps and GPS for mobile phones allow one to scan an area and receive information about nearby restaurants, hotels, pharmacies and museums). On the other end, VWs are also getting closer to photorealistic environments and can even import aspects of the physical world through features such as Google Mars and geographic information systems. The merging of the physical with the virtual has great potential for long duration space missions. Through wireless transmission, augmented reality and novel, embodied ways of communication in VWs, newer forms of interaction with fellow humans and our environment abound.

VWs can play a crucial role in acclimating us to other versions and conditions of reality that exist in the extra-terrestrial. Moreover, since they also can be mined for data, they can be used to monitor progress, both in pre-mission training, and throughout future missions where VW applications are incorporated, In mining data, benchmarks can be established for physical, psychophysical and psychological strengths that are required to exist and operate in extra-terrestrial realities. VWs can be personalized and enhanced with Embodied Conversation Virtual Agents and Intelligent Tutoring Systems, as well as with characters that can aid as countermeasures for negative factors associated with isolation and separation. As humans, we can also learn about the self in relation to other human by connecting to these artificial intelligence systems. Finally, through the broad range of scenarios possible through VWs, ASCANS can learn to prepare and strengthen their minds for complex problems and adverse conditions that they will encounter in long duration space flight. They can learn how to deal with fellow human beings, and optimize decision-making and leadership skills. They can find relaxation, renewal and comfort within the simulacrum of virtual worlds, as well as connect to their loved ones, and learn to acclimate to and deal autonomously with the communication delays they will encounter. As we hope to have shown in the above examples and recommendations, the potential of VWs for astronauts, flight and ground crew and overall mission success cannot be overlooked as NASA proceeds into the future. The benefits of the richness of virtual worlds and their realization as embodied Cyberspace, constitute a crucial bridge to ensuring successful long duration space missions, those future voyages that will serve to evolve all humankind and bring us better understanding of our place in the universe.

Glossary of Terms

Avatar – literally, the "embodiment" of someone in a virtual space.

CMS – Course Management Systems, which are online systems that are used to track learning that are currently used for distance learning

ECA – Embodied Conversational Agents are virtual agents that can respond and interact intelligently with a person through natural language processing, which makes semantic inferences on what is being said by that person by matching their language to various scenarios

IA – Intelligent Agents that act autonomously to respond to a person using Natural Language Processing and thus appear as artificially intelligent

ITS – Intelligent Tutoring System uses the input of a person's responses on a task, gauges their performance on the task, and asks follow-up questions that provide that person with post-task reflection. ITS can provide important mentoring opportunities for students.

NLP – Natural Language Processing is the use of artificial intelligence to parse either voice or text input into understanding of what is being conveyed in the input and responding to the input through computer artificial intelligence output

VE – Virtual Environments are spatial virtual reality applications

Virtual Humans – Intelligent, embodied conversational agents that can converse with a person for a variety of purposes and create a more complex simulation.

VR – Virtual Reality – early versions of virtuality where reality is simulated in an applied manner

VWs – Virtual Worlds are social versions of virtual reality environments and use computer-mediated communication such as instant messaging, chat, and avatar-to-avatar interactions

References

Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (2000). *Embodied Conversational Agents*. Cambridge, MA: MIT Press.

Fisher, S. S., McGreevy, M., Humphries, J., & Robinett, W. (1986). Virtual Environment Display System, *ACM 1986 Workshop on 3D Interactive Graphics* (pp. 77-87). New York: ACM Press.

Gibson, W. (1984). *Neuromancer*. New York, NY: Ace.

Gibson, W. (1986). *Burning Chrome*. New York, NY: Ace.

Hall, E.T. (1980). *Beyond culture*. New York: Anchor.

Hall, E.T. (1982). *The hidden dimension*. New York: Doubleday.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage.

Marsh, J. (2010). Young children's play in virtual worlds. *Journal of early childhood research 7*(3), 1-17. Thousand Oaks, CA: Sage.

Morie, J.F., Verhulsdonck, G. & Lauria, R. (2010a). *Countermeasure technologies for long duration spaceflight: Using virtual world technologies to enhance physical, psychological & team performance*. [Report]. Literature Review. Houston, TX: BHP/Wyle Laboratories.

Morie, J.F., Verhulsdonck, G. & Lauria, R. (2010b). *Applying countermeasure technologies to augment current training protocols for long duration space flight: Using virtual worlds to enhance team performance, leadership skills, conflict management and communication style*. [Report]. Operational assessment. Houston, TX: BHP/Wyle Laboratories.

Morie, J.F., Lauria, R., & Verhulsdonck, G. (2010c). *Applying countermeasure technologies to augment current training protocols for long duration space flight: Using virtual worlds to enhance team performance, resiliency, and proficiency skills.* [Report]. Expert panel interviews. Houston, TX: BHP/Wyle Laboratories.

Prensky, M. (2005). *Digital-Game Based Learning*. New York: McGraw-Hill.

Stephenson, N. (1992). *Snow crash*. New York: Bantam.

Wagner Au, J. (2009). *Other World Notes: 1 in 8 Americans Virtual World Users. [online]* Retrieved, August 27th, 2010 from: http://nwn.blogs.com/nwn/2009/07/other-world-notes-5.html

Yee, N., Bailenson, J.N., & Ducheneaut, N. (2009). The Proteus Effect: Implications of Transformed Digital Self-Representation on Online and Offline Behavior. *Communication Research*, 36, 285-312.

FINAL REPORT

Investigating the Influence of Personality on Astronaut Career Performance

Prepared by
David Musson, MD, PhD
Contractor

For
Kathryn E. Keeton, PHD
Senior Scientist
Behavioral Health and Performance (BHP) Research Element
EASI/Wyle NASA BioAstronautics Contract

September 30, 2010

Contents

1. Overview

1.1. This report presents an analysis of an existing of astronaut psychological trait dataset and the relationship between those data and publicly available metrics of astronaut career performance. This project is funded by a contractual agreement between the author and the Behavioral Health and Performance (BHP) Research Element at EASI/Wyle and the National Aeronautics and Space Administration (NASA). This work has been undertaken for the purpose of informing future selection strategies for astronaut applicants, and to create a better understanding the relationship between individual psychological characteristics and the job of being and astronaut. As per long standing agreements between participants and the investigators who collected these data, no direct sharing of any data that may be individually identifiable shall be made between the author of this report and NASA or any of its agencies.

2. Understanding the personal characteristics data set

2.1. This investigation of the relationship between astronaut personal characteristics and astronaut career performance involves the reassembly of two archived data sets. Two distinct sources of data were collected between 1989 and 1995. The first of these was derived from the original study of astronaut personality and performance conducted by Helmreich, Rose, et al – begun in 1988 with results published in 1993 and 1994 (Rose, Fogg, Helmreich, & McFadden, 1994). This study involved testing the population of 65 active duty astronauts in 1989. Of the 89 active astronauts who were serving during this period, a total of 65 consented to and participated in this original data collection. The papers resulting from this study were the first, and to-date the only published formal investigation into personality and performance in this population. These original studies investigated the relationship between personality, as assessed by trait testing batteries, performance, as assessed by a series of peer and supervisory ratings on multiple parameters. Findings were modest in these studies, but suggested a possible link between interpersonal orientation and peer/supervisory rated performance (McFadden, Helmreich, Rose, & Fogg, 1994; Rose, Fogg, Helmreich, & McFadden, 1994; Rose, Helmreich, Fogg, & McFadden, 1993).

2.2. The second data set included in this analysis was collected during the NASA astronaut recruitment campaigns that were held in the 6 years following the above study (1989-1995). These data were collected from the 259 astronaut applicants who participated in the NASA final-stage astronaut selection process carried out during this period. Of the 259 astronaut applicants on whom data were collected during final selection, a total of 63 (12 female, 51 male) were eventually successful in their application to become astronauts. Formal analyses of these data have been previously presented (Musson, Sandal, & Helmreich, 2004).

2.3. It should be noted that subtle score differences on personality measures between those already selected and those undergoing selection have been described (Sandal, Musson, Helmreich, & Gravdal, 2004). The theorized difference between these two data sets is due to the tendency of job applicants to present themselves in an especially positive light due to their participation in the astronaut selection process at the time of testing. Even though the tests were presented to astronaut applicants as research measures, it is likely that either a halo of positive self-presentation or a mistrust that the testing results would actually be used in selection, despite reassurances to the contrary.

2.4. Combining the two data sources described above puts the total number of astronauts in this current analysis at N=65+63=128. It is recognized that the different conditions under which these two populations were assessed did in fact lead to small, but systematic biases in the scores in personality traits. Records indicate that consent was obtained from subjects in both data collections for the use of these data for ongoing selection and performance studies. Guarantees of both anonymity and arms-length separation of these data from NASA management were assured as a condition of participation, and these promises have been kept to date by the Principal Investigator (Helmreich) and his associates (ie, the author of this report).

3. *Personal characteristics as predictors of performance*

3.1. The overall aim of the analysis presented in this report is to examine the relationship between individual factors (i.e. predictors), identifiable at the time of selection, to career activity (i.e. performance) as an astronaut. Funding for these analyses did not include the collection of new data on astronaut performance beyond those data that are already available in the public domain. Unlike the original studies into personality that were described above and which looked at peer/supervisory assessments or perceived competence, interpersonal skill, and social compatibility for a proposed long duration flight, the present analysis relies on metrics derived from flight assignment and career promotion. As will be discussed later in this report, a lack of a clear theoretical alignment between predictors and outcome is less likely to demonstrate significant results in the analysis, compared to studies where those two factors are more theoretically align. For a detailed discussion of this concept, the reader is referred to Hogan and Holland, 2003 (Hogan & Holland, 2003).

3.2. These predictors available for analysis in this report fall under two broad categories: **Demographic** predictors and psychological **Trait** predictors.

3.3. **Demographic predictors**. Demographic variables race are easily identifiable from public record, and include: Age, Gender, Military service vs Non-Military service, Pilot vs Mission Specialist, Academic background (Engineer, Medical Doctor, Scientist), Educational Level. In general, these are not of particular interest in the present analysis except as potential explanations of variance that are not attributable to personality traits. A full analysis of these predictors is beyond the scope of this report, though known interactions between personality trait scores and gender is of some interest.

3.4. **Trait predictors**. Trait variables, assessed in the two astronaut population grps groups and for which data exist on each participant are appended in some detail in **Appendix A.** The measures used to collect these data represented those instruments that had been used in aviation performance studies by the principal investigator of the original astronaut performance study conducted in 1989 (Helmreich), along with adaptations of additional scales that were emerging in the literature at the time of the original study.

3.5. Measures existent in the current data sets include the Instrumentality and Expressivity scales of the Personal Attributes Questionnaire and the Extended Personal Attributes Questionnaire (Helmreich, Spence, & Wilhelm, 1981). Also included were scales from the Jenkins Activity Survey (Jenkins, Zyzanski, & Rosenman, 1971), the Personality Research Form (Jackson) (Jackson, 1997), the Work and Family Orientation Questionnaire (Helmreich, 1978), and modified versions of the NEO-Five Factor Inventory (NEO-FFI) questionnaire (Costa & McCrae, 1992), which was a relatively new measure at the time of assessment, and which has since emerged to become the dominant trait measure currently in use in the field of organizational psychology some 20 years later. These 5, twelve-item scales of the NEO-FFI were reduced to 5, eight-item scales in order to accommodate mandated restrictions in testing astronaut participants. Answer keys were also modified somewhat to accommodate the larger battery of tests that were administered to the astronauts at the time of testing. Formal comparisons have been made between the full NEO-FFI and the modified versions used in these assessments (Musson & Helmreich, 2003), and while some diminution in scale quality was evident, the scales still meet adequate levels of reliability for research purposes. Numbers of astronauts who have completed each of the scales of the abovementioned instruments are provided in **Appendix B** of this report.

3.6. There has also been discussion over the years of *High Instrumental + High Expressivity*, *High Instrumental + Low Expressivity*, and *Low Instrumental* clusters in the astronaut (and other) populations. These have also been categorized as the *Right Stuff*, *Wrong Stuff*, and *No Stuff* clusters (Chidester, Foushee, & Jensen, 1991). This information is included for completeness, and may be examined at a later date. Although conceptually satisfying, the current plan is not to include this cluster-based analysis in this data as such analysis have not proved to be a mainstream means of analyzing personality-performance relationships since the model was first introduced in the 1980s.

3.7. A number of additional scalar variables are listed in Appendix A, such as those of the Personality Research Form and some scales from the Work and Family Orientation questionnaire. These variables are also not planned to be included in the analysis in this study, as a significant body of literature does not presently support the use of those scales in vocational performance research. Variables selected for analysis include the following, derived from the Helmreich inventory:

3.7.1. Instrumentality
3.7.2. Expressivity
3.7.3. Negative instrumentality
3.7.4. Verbal aggressiveness
3.7.5. Negative communion

3.7.6. Mastery
3.7.7. Work orientation
3.7.8. Competitiveness

3.7.9. Impatience/irritability
3.7.10. Achievement strivings

3.8. And, derived from the NEO-FFI are the following Big Five scales:

3.8.1. Neuroticism

3.8.2. Extraversion

3.8.3. Openness to new experience

3.8.4. Agreeableness

3.8.5. Conscientiousness

4. Outcome variables – quantifying career performance

4.1. Rating astronaut performance is extremely challenging. This is an unusually high-performance population, and identifying criteria that may serve to distinguish levels of career performance is difficult. This challenge has been noted previously in previous publications and workshops (McFadden, Helmreich, Rose, & Fogg, 1994; Musson & Helmreich, 2005). NASA does not publicly discuss performance differences between individual astronauts, and there are no universally agreed-upon metrics of performance in this population. Furthermore, performance evaluations are sensitive topics under ideal circumstance, and in this highly competent, publicly scrutinized population care must be taken to protect individuals from embarrassment, criticism, or negative career actions.

4.2. Peer and supervisory evaluations: Often, job performance measures include supervisory and peer ratings, based on the assumption that peers and supervisors are in a position to make both specific and global assessments of individuals' performance and to make comparisons with other workers. In the context of this project, this would include peer assessments of astronauts by other astronauts, supervisory assessments of astronauts by individuals in positions of authority within the NASA community, or ratings of astronauts by non-astronaut co-workers. Neither funding nor opportunity exists at this stage to pursue this avenue, though it could be most valuable to pursue down the road. It should be noted that the original astronaut performance studies conducted by Rose and Helmreich failed to find statistically solid predictors of peer and supervisory ratings within this population (Rose, Fogg, Helmreich, & McFadden, 1994), though inferences were made on the relationship between traits related to interpersonal competence and perceived desirability for cohabitation, which makes intuitive sense. One of the main limitations of those analyses was the overall N being only 65, despite high levels of participation from the astronaut corps.

4.3. In the absence of a funded project to conduct peer or supervisory assessments, publicly available parameters of astronaut performance have been identified in this project as the most appropriate initial measure of job performance in this population. The public record include multiple measures that may be interpreted as measuring success/failure at both the mission and career level. Presumably, since superior performers are identified through peer, supervisory and training evaluation (both formal and informal), early assignment to flights, assignment to high responsibility roles (eg, mission commander or extravehicular activities (EVA)), frequent flight assignments and career longevity can all be theorized to serve as proxies for high internal assessments of competence and effectiveness within the astronaut corps. Possible outcome variables, based on observable career performance includes:

4.3.1. Binary (Yes/No) outcome variables

4.3.1.1. Assignment of command positions on spaceflights

4.3.1.2. Assignment to CapCom position

4.3.1.3. Leadership positions within the astronaut office (Chief or Deputy)

4.3.2. Scalar outcome variables:

4.3.2.1. Time delay to first flight assignment

4.3.2.2. Number of flight assignments

4.3.2.3. Number of EVA assignments over an astronaut's career

4.4. Command positions and CapCom positions presumably represent some measure of perceived competence. Election to leadership assignment would appear to be a meaningful measure of peer assessment of competence. Numbers of both Chief and Deputy Chief of the Astronaut Office are relatively low, so a combined statistic would appear to be most useful. Ten individuals in this dataset have occupied one or both of these leadership positions.

4.5. Time delay between recruitment and first flight is a relatively simple calculation and has been calculated for each astronaut in the data set. Astronauts more rapidly assigned to a flight following training may indicate perceived higher levels of competence. This viewpoint is admittedly one of conjecture, and certainly many factors may affect such assignment. Still, it is a relatively easy variable to calculate and it warrants exploration. Similarly, number of flights and number of EVAs for each astronaut can be calculated, and will be used for analysis. The decision was made not to consider either total flight time or number of hours of EVA, as these would appear to provide little additional information over and above numbers of flight assignments and number of EVAs.

5. Analyses of the data

5.1. The analytic data set

5.1.1. The above discussion of predictors and outcomes refers to the raw data set used in these analyses. Of the existing data, exclusion of a modest number of individuals is warranted.

5.1.2. A small number of those represented in this data set began their careers well before the vast majority of other respondents. Comparison of the career performance of these individuals with the rest of the corps is not a relevant comparison due to significant differences in career opportunities and highly differing opportunities for operational experience. These individuals will be excluded form the analysis.

5.1.3. Also, and most unfortunately, a number of individuals in this data set are deceased. These astronauts never had the opportunity to live out their career to the extent their colleagues where able to do – as such, inclusion of these individuals in these analyses does not fit with predictors of long-term performance in the corps.

5.1.4. Specific numbers in each of these two categories are not provided in order to protect the anonymity of the respondents, consistent with the agreements made at the time of data collection. The data set remaining after the exclusion of these individuals and that will be used for subsequent analysis is as follows:

Demographics by gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	FEMALE	22	18.5	18.5	18.5
	MALE	97	81.5	81.5	100.0
	Total	119	100.0	100.0	

5.1.5. On the binary set of performance measures, the following tables represent the metrics within the data set:

Command position

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No Command	59	49.6	49.6	49.6
	Command	60	50.4	50.4	100.0
	Total	119	100.0	100.0	

Capcom

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Capcom	17	14.3	14.3	14.3
	No Capcom	102	85.7	85.7	100.0
	Total	119	100.0	100.0	

Leadership position?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Chief or Deputy	10	8.4	8.4	8.4
	Neither	109	91.6	91.6	100.0
	Total	119	100.0	100.0	

5.1.6. For the Scalar outcome variables, the following values (mean, standard deviation) represent the performance characteristics in the data set used for these analyses. Ranges are not provided, as these may allow some identification of specific participants.

5.1.6.1. Number of Missions: Mean = 3.40 Std dev = 1.284

5.1.6.2. Time to first flight: Mean = 4.81 yrs, Std dev = 2.168

5.1.6.3. Number of EVAs: Mean = 1.35 Std dev = 2.200

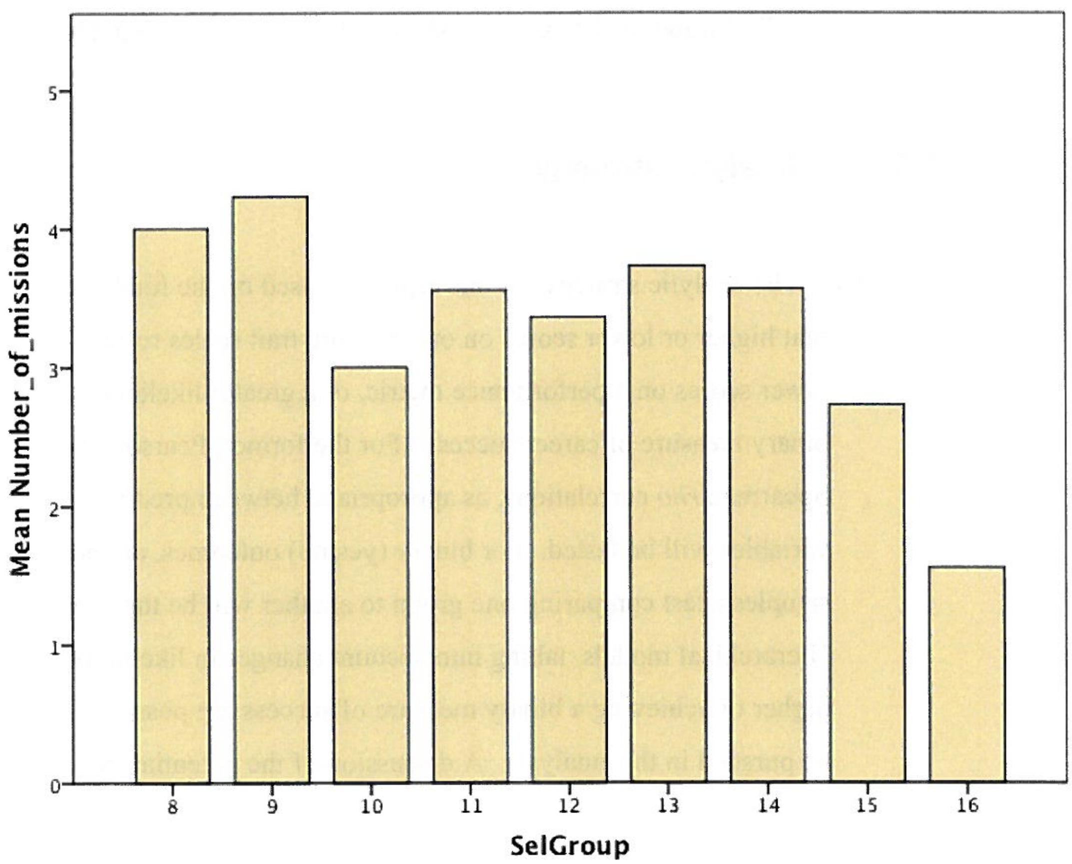
5.2. Analytic strategy

5.2.1. The analytic strategy for this report is based on the fundamental model that higher or lower scores on one or more trait scales results in higher or lower scores on a performance metric, or a greater likelihood of attaining a binary measure of career success. For the former, Pearson product or Spearman *rho* correlations, as appropriate, between predictor and outcome variables will be tested. For binary (yes/no) outcomes, an independent samples t-test comparing one group to another will be the preferred method. Hierarchical models, taking into account changes in likelihood of scoring higher or achieving a binary measure of success are possible, though will not be pursued in this analysis. A discussion of the potential role for such analyses will be explored in the discussion section at the end of this report.

For all analyses, an alpha of 0.05 will be used as the statistical cut-off for determination of significance. It is recognized that failure to correct for the rather large number of comparisons conducted will result in an increased likelihood of finding correlations that are actually spurious, but the exploratory nature of these analyses and the unique and limited nature of these data suggest that such an approach is warranted. These deficiencies are acknowledged.

5.3. Demographic factors and astronaut career performance:

5.3.1. **Selection year and number of flight assignments.** The relationship between selection year and number of flight assignments is shown graphically below.



5.3.2. The above graph demonstrates the relationship between selection group (or class) and the mean number of flight assignments per class. The correlation between selection class and flight assignment, as suggested by the above graph, is not insignificant, and the Spearman's rho correlation coefficient between group number and mean flights is 0.320, $p<0.001$. This means that astronauts selected in successive years were likely assigned to a fewer number of flights. Some of this will be due to the fact that many of these later astronauts have more years of productivity and theoretically more possible flight assignments ahead of them, but it may also reflect real decreases in available flights or lower chances of flight assignments relative to their more experienced peers

5.3.3. Gender and career performance: It is clear that the number of female astronauts is significantly less than the number of male astronauts in the current astronaut population compared to the general non-astroanut population. In our predictor/performance data set of 129 individual astronauts, 23 are female and 106 are male (17.8% and 82.2%). Among female astronauts, any significant differences between males and female numbers in terms of flight assignments, leadership positions, and EVA assignment may account for variance in performance outcome measures that would eclipse trait predictors. While a full analysis of gender and its role in selection and astronaut careers is well beyond the scope of this project, the following are observed.

5.3.4. In the post-Apollo era of U.S. spaceflight (ie, when females were eligible to selected as astroauts), a total of 17 individuals have been assigned as Chief or Deputy Chief positions in the Astronaut Office. Of that number, 4 have been female and 13 were male (23.5% vs 76.5%), which closely approximates the distribution within the population. Only 1 female has been Chief during this period, where as 6 males have held the post (14% vs

86%). The current dataset does not have adequate representation from leadership roles to justify controlling for gender as a potentially influencing variable of this parameter.

5.3.5. In the dataset being used for this analysis, the relationship between gender and flight assignment suggests no impact of gender on frequency of flight assignment. Female astronauts average 3.09 flights per career whereas males average 3.33 flights per career. This difference is not significant.

5.3.6. The same is not true for EVAs, however. Only 7 of 23 females in the data set have performed EVAs (30%), where as 53 of 106 males performed EVAs (50%). This result is statistically significant ($t = -3.320$, $p=0.001$, equal variances not assumed). Males who performed EVAs were also likely to perform additional EVAs, whereas this was not the case for females.

6. Trait predictors and performance:

6.1. Among the large number of personality variables in the existing database, the decision was made to focus on two broad subsets, based on support from the literature. The first set is the family of trait measures that is described in the astronaut performance literature as the Helmreich Personal Characteristics inventory. As a reminder, these measures include the following:

6.1.1. Instrumentality
6.1.2. Expressivity
6.1.3. Negative instrumentality
6.1.4. Verbal aggressiveness
6.1.5. Negative communion

6.1.6. Mastery

6.1.7. Work orientation

6.1.8. Competitiveness

6.1.9. Impatience/irritability

6.1.10. Achievement strivings

6.2. As mentioned elsewhere, the first 5 are draw from the Personal Attributes Questionnaire (Helmreich, Spence, & Wilhelm, 1981), the next three from the Work and Family Orientation Questionnaire(Helmreich, 1978), and the last two from the Jenkins Activity Survey (Jenkins, Zyzanski, & Rosenman, 1971). These 11 scales were chosen because, among the non-Big Five (see below) scales that were included in this battery, they have the strongest justification for assuming a personality-performance relationship based on previous research (Helmreich, Spence, Beane, Lcker, & Matthews, 1980; Helmreich, Spence, & Pred, 1988).

6.3. Five additional personality scales are included in the analytic strategy for this project:

6.3.1. Neuroticism

6.3.2. Extraversion/Introversion

6.3.3. Openness to new experience

6.3.4. Agreeableness

6.3.5. Conscientiousness

6.4. These are variants of the scales designed to assess the well known family of traits that are typically referred to as the Big Five (McCrae, Costa, Pervin, & John, 1999). This scales are derived from the principal insrument used to assess the Big 5, the NEO-FFI developed by Costa and McCrae (Costa & McCrae, 1992). The Big Five have been the mst popular scales for investigations of personality and performance over the last 15 years (Fruyt & Mervielde, 1999; Judge, Higgins, Thoresen, & Barrick, 1999; Mount & Barrick, 1991). It must be noted that these scales were modified at the time of testing in both scale length and answer key, so as to better integrate with the testing battery that was used for the original data collection. An analysis of these scales has been conducted previously, and acceptable relationships to the original scales has been demonstrated (Musson & Helmreich, 2003).

6.5. Scores did not demonstrate significant differences between genders in this population, and are presented below with normative scores drawn from test populations for comparison.

	Astronauts		Female normative		Male normative	
	Mean	SD	Mean	SD	Mean	SD
Instrumentality	26.65	2.89	20.61	4.52	21.35	4.83
Expressivity	22.69	3.57	24.74	3.74	23.27	3.97
Negative instrumentality	10.01	4.39	11.14	4.67	13.35	4.49
Verbal aggressiveness	4.24	2.43	6.54	2.88	5.29	2.87
Negative communion	5.24	2.05	7.15	2.07	6.35	2.32
Mastery	22.99	3.59	18.89	4.77	19.12	4.38
Work orientation	22.53	1.60	20.62	3.06	19.07	3.73
Competitiveness	12.17	3.68	13.48	3.86	14.13	4.16
Impatience/irritability	9.84	2.84	10.40	3.88	10.70	3.51
Achievement strivings	18.28	3.14	14.38	3.79	13.71	4.49

6.6. Data from the existing astronaut personality database is presented above, along with normative data broken down by gender. Significant gender differences are not present in the astronaut data, suggesting a gender-independent astronaut profile. It can be seen in the table above that astronauts have higher than usual scores on Instrumentality and related scales, slightly lower scores on Expressivity, and lower scores on verbal aggressiveness and negative communion.

	Astronauts		Female normative		Male normative	
	Mean	SD	Mean	SD	Mean	SD
Neuroticism	6.05	3.40	15.13	5.89	14.24	5.78
Extraversion	23.12	3.46	22.83	4.69	20.97	4.57
Openness	19.73	4.63	18.02	4.90	19.59	4.80
Agreeableness	24.71	3.21	22.35	4.07	20.22	4.16
Conscientiousness	26.66	3.20	21.69	5.04	19.28	5.65

Normative values are taken from Musson & Helmreich, 2003.

6.7. Big 5 data are presented above, along with normative data. Again, systematic gender differences are not present in the astronaut data. It can be seen in the table above that the astronauts, as a population, have unusually low scores on Neuroticism, are mildly high on Extraversion, and particularly high on Agreeableness and Conscientiousness. Particularly with Conscientiousness, a well established predictor of performance (Mount & Barrick, 1991), ceiling effects from uniformly high scores and relatively little individual variation within the population may limit the usefulness of the proposed analysis.

6.8. Personality traits and command positions

6.8.1. The relationship between personality and assignment to in-flight command positions was explored. Of the 15 personality variables considered, only **Openness to New Experience** was significantly related, with command-assigned individuals scoring slightly lower than their non-command counterparts (18.60 vs 20.96), $t = 2.52$, $p=0.013$.

6.9. Personality traits and Capcom Assignment

6.9.1. The relationship between personality and assignment to Capcom positions was explored. Of the 15 personality variables considered, **none** showed a significant difference between those selected and those who were not.

6.10. Personality traits and administrative leadership (Chief/Deputy AO)

6.10.1. Independent samples t-tests were conducted comparing mean scale scores of those astronauts who have held leadership positions, i.e. those who have held either Chief or Deputy Chief, Astronaut Office ($N = 10$), and those who have not ($N = 114$). Only one trait predictor variable showed a difference in mean scores with a significance of less than $p < 0.05$ – that being **Work Orientation**, ($t = -2.108$, $df = 122.000$, $p = 0.037$). Work Orientation has been defined as, "reflecting a general desire to work hard." Interestingly, those astronauts who have held such leadership positions scored less than those who have not (mean score 21.7143 vs 22.5128) on this measure.

6.11. Personality measures and time to first flight

6.11.1. Time to first flight ranged from 2 to more than 9 years, with a mean of 5.04 yrs (SD = 2.636). Of the 15 personality scales, **Neuroticism** was positively correlated with time to first flight (r = 0.2.15, p = 0.042).

6.12. Personality measures and number of flight assignments

6.12.1. Number of flight assignments ranged from 0 to 7, with a mean of 3.29 flights per astronaut (SD = 1.324). No significant relationship was found between number of flight assignments and any of the trait predictors.

6.13. Personality measures and number of EVAs

6.13.1. Number of EVAs range from 0 to more than 7, with a mean of 1.29 per astronaut (SD= 2.141). As expected, number of EVAs correlated to some extent with number of flight assignments (r = 0.221, p = 0.012). However, EVAs were also correlated negatively with **Openness to new experience**, with r = -0.239, p = 0.020. No other correlations were identified.

7. Summary of analysis

7.1. Data analyses showed weak to little correlation between demographic predictors and flight assignment, EVA assignment, and CapCom assignment with the exception of gender and EVA assignment. Female astronauts were less likely than their male counterparts to be assigned EVAs, at least among our participants. No conclusions can be drawn in this report regarding the larger astronaut population where this difference may or may not persist, as those data were not analyzed. Selection year did indeed predict likelihood of fight assignment, though it can not been determined whether this is due solely to career longevity, or whether this difference is likely to remain at the end of the careers of those who were more recently recruited.

7.2. The tendency for those in **leadership positions** to score slightly less on **Work Orientation** compared to non-leadership colleagues is difficult to explain, though it may fit with some lay theories held by workers regarding management in general. This finding will likely be popular with non-management individuals, and less popular with or simply discounted by management.

7.3. Since opportunities for flight assignment vary significantly from year to year, it is difficult to compare astronauts recruited in one year to those recruited in another year (see the following section). Regardless, some modest correlations were identified. With respect to **time to first flight**, higher scores on **Neuroticism** were related to delayed first flight. This is consistent with a theory that low levels of Neuroticism may be associated with perceptions of appropriateness for flight assignment. This also fits with traditional perceptions of test pilots and astronauts being calm-mannered individuals adept at handling the emotionally provocative environment of high-risk flight. The correlation between low scores on **Openness to new experience** and **career EVAs** is more difficult to explain, and though possibly spurious, may warrant further study.

8. *Suggestions for next steps*

8.1. The modest though real relationship between selection class and flight assignment confirms that some small portion of the variance in analysis of predictors of numbers of flights assigned is due to this effect of selection class. A more complex hierarchical analysis of the relationship between personality predictors and this outcome may be of some value. However, an inspection of the correlation matrix of personality scales to number of flights shows no relationships that are nearing significance, with the exception of a weak and statistically insignificant correlation with Extraversion ($r = -0.191$, $p=0.065$). If this finding were to become statistically significant through hierarchal modelling, it would likely be one of low magnitude and questionable interpretation.

8.2. A more meaningful, though more extensive next step would involve funding a more appropriate model of outcome measures. Personality, in this highly selected population is unlikely to have significant impact on the measures considered in this study (and the analyses herein confirms this statement). All selected astronauts have impressive records of achievement prior to selection, which is a fact consistent with their uniformly high scores on achievement and motivational traits presented in this report. In such a population, many factors over and above personality are likely to determine the blunt, insensitive metrics that were used in these analysis – factors such as previous experience and technologically-specific knowledge certainly contribute to flight selection, EVA selection, and Capcom assignment. Where personality is more likely to have an impact is in less technical metrics. Appropriateness for long-duration cohabitation, crew selection for multicultural missions, and other socially oriented factors are likely to be based largely on personal characteristics. Peer assessments of desirability on such scales should more closely align with personality traits, and an analysis should focus on whether standardized personality measures are able to identify those individuals destined to be peer-identified as more desirable. Traits more likely to predict such assessments would include Agreeableness, Expressivity, and other traits related to social interaction. Among those predictors, the astronaut population shows more similarity to the general population, and variability is greater than on achievement related traits in this population. Such a study would require revisiting methodology, and would most certainly require new ethics reviews, informed consent, and high degrees of collaboration from the active and retired astronaut population.

9. References

Chidester, T. R., Foushee, H. C., & Jensen, R. (1991). Leader personality and crew effectiveness: a few mission simulation experiment. In *Proceedings of the 5th International Aviation Psychology Symposium Vol. II.* Columbus, OH: Ohio State University.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R Professional Manual.* Odessa FL: Psychological Assessment Resources, Inc.

Fruyt, F. D., & Mervielde, I. (1999). Raisec Types and Big Five Traits as Predictors of Employment Status and Nature of Employment. *Personnel Psychology, 52*, 701-727.

Helmreich, R. L. (1978). The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes towards family and career. *JSAS Catalog of Selected Documents in Psychology, 8*(35), MS 1677.

Helmreich, R. L., Spence, J. T., Beane, W. E., Lcker, G. W., & Matthews, K. A. (1980). Making it in academic psychology: Demographic and personality correlates of attainment. *Journal of Personality and Social Psychology, 39*, 896-908.

Helmreich, R. L., Spence, J. T., & Pred, R. S. (1988). Making it without losing it: Type A, achievement motivation, and scientific attainment revisited. *Personality and Social Psychology Bulletin, 14*(3), 495-504.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles, 7*, 1097-1108.

Hogan, J., & Holland, B. (2003). Using Theory to Evaluate Personality and Job Performance Relations: A Socioanalytic Perspective. *Journal of Applied Psychology 88*(1), 100-112.

Jackson, D. N. (1997). *Jackson Personality Research Form.* Port Huron, MI: Sigma Assessment Systems.

Jenkins, C. D., Zyzanski, S. J., & Rosenman, R. H. (1971). Progress toward validation of a computer-scored test for the Type A coronary prone behaviour pattern. *Psychological Medicine, 33*, 193-202.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*, 621-652.

McCrae, R. R., Costa, P. T., Pervin, L. A., & John, O. P. (1999). A Five Factor Theory of Personality. In *Handbook of Personality: Theory and Research* (Vol. 1 pp. 139-153). New York, NY: The Guilford Press.

McFadden, T. J., Helmreich, R. L., Rose, R. M., & Fogg, L. F. (1994). Predicting astronaut effectiveness: a multivariate approach. *Aviation Space & Environmental Medicine, 65*(10 Pt 1), 904-909.

Mount, M. R., & Barrick, M. K. (1991). The big five personality dimensions and job performance: a meta analytic review. *Personnel Psychology, 44*(1), 1-26.

Musson, D. M., & Helmreich, R. L. (2003). *Psychometric Properties of 8-Item Measures of the Big Five Derived from the NEO Five Factor Inventory (NEO-FFI) Technical Report TR-2003-01.* Austin, Texas: University of Texas Human Factors Research Project.

Musson, D. M., & Helmreich, R. L. (2005). Long-term data sets in space behavioral research. *Aviation Space & Environmental Medicine, 76 suppl*, B116-125.

Musson, D. M., Sandal, G. M., & Helmreich, R. L. (2004). Personality characteristics and trait clusters in final stage astronaut selection. *Aviation Space & Environmental Medicine, 75*, 342-349.

Rose, R. M., Fogg, L. F., Helmreich, R. L., & McFadden, T. J. (1994). Psychological predictors of astronaut effectiveness. *Aviation Space & Environmental Medicine, 65*(10 Pt 1), 910-915.

Rose, R. M., Helmreich, R. L., Fogg, L., & McFadden, T. J. (1993). Assessments of astronaut effectiveness. *Aviation Space & Environmental Medicine, 64*(9 Pt 1), 789-794.

Sandal, G. M., Musson, D. M., Helmreich, R. L., & Gravdal, L. (2004). Social desirability bias in personality testing; implications for Astronaut selection. In *Proceedings of the 55th International Astronautical Congress 2004.* Vancouver, Canada: International Astronautical Federation.

APPENDIX A – Personal Characteristic Scales

Astronaut version of the Personal Characteristics Inventory (astrPCI) Subscales and their test origins.

Personal Attributes Questionnaire PAQ/ Extended Personal Attributes Questionnaire (EPAQ) – Helmreich and Spence

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles, 7*, 1097-1108.

1. Positive expressivity
2. Instrumentality
3. Bipolar instrumentality
4. Negative instrumentality
5. Negative expressivity - verbal aggression
6. Negative expressivity - negative communion

Work and Family Orientation Questionnaire (WOFO) – Helmreich and Spence

Helmreich, R. L. (1978). The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes towards family and career. *JSAS Catalog of Selected Documents in Psychology, 8*(35), MS 1677.

1. Mastery
2. Work orientation
3. Competitiveness
4. Work involvement
5. Joy in work
6. Job involvement
7. Perfectionism
8. Driven

Jenkins Activity Survey (JAS) – Jenkins

Jenkins, C. D., Zyzanski, S. J., & Rosenman, R. H. (1971). Progress toward validation of a computer-scored test for the Type A coronary prone behaviour pattern. *Psychological Medicine, 33*, 193-202.

1. Impatience/irritability
2. Achievement striving

Personality Research Form (PRF) – Jackson

Jackson, D. N. (1997). Jackson Personality Research Form. Port Huron, MI: Sigma Assessment Systems.

1. Affiliation
2. Dominance
3. Endurance
4. Impulsivity
5. Succorance
6. Alienation
7. Vigor

Neo Five Factor Inventory (NEO-FFI) – Costa and McCrae – Modified

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R Professional Manual.* Odessa FL: Psychological Assessment Resources, Inc.

1. Neuroticism
2. Extraversion
3. Openness
4. Agreeableness
5. Conscientiousness

Additional scales from the Spence Workaholic research measures

1. Joy In Work (JOYINWORK)
2. Job Involvement (JOBINVOL)
3. Perfectionism (PERFISM)
4. Time Commitment (TIMECOMM)
5. Work Driven (DRIVE)

APPENDIX B – Number of Respondents

Number of astronauts completed each of the astro-PCI personality trait scales

PAQ/EPAQ	N
Instrumentality	128
Expressivity	128
Bipolar Instrumentality	127
Negative Instrumentality	129
Verbal Aggressiveness	129
Negative Communion	128

WOFO	N
Mastery	129
Work Orientation	128
Competitiveness	129

JENKINS ACTIVITY SURVEY	N
Impatience/Irritability	128
Achievement Strivings	128

NEO-FFI (Modified)	N
Pci-Neuro	104
Pci-Extra	104
Pci-Open	104
Pci-Agree	104
Pci-Consc	104

PRF	N
Affiliation	41
Dominance	63
Endurance	63
Impulsiveness	63
Succorance	63
Social Alienation	63
Physical Vigor	41

Spence Work Scales	N
Work Driven	60
Joy In Work	60
Job Involvement	60
Perfectionism	38
Time Commitment	60

APPENDIX C – Right Stuff Cluster Definitions

Personal Characteristics Inventory clusters defined
Right Stuff, Wrong Stuff, No Stuff

Cluster 1 (positive instrumental/expressive)

(Right Stuff)

Above average in:

Instrumentality (EPAQ)
Expressivity (EPAQ)
Mastery (WOFO)
Work (WOFO)

Below average in:

Negative Instrumentality (EPAQ)
Verbal aggressiveness (EPAQ)

Cluster 2 (negative instrumental)

(Wrong Stuff)

Above average in:

Instrumentality (EPAQ)
Negative Instrumentality (EPAQ)
Verbal Aggressiveness (EPAQ)
Work (WOFO)
Mastery (WOFO)
Competitiveness (WOFO)

Below average in:

Expressivity (EPAQ)

Cluster 3 Low motivation (No Stuff)

Low scores in the following scales:

Instrumentality (EPAQ)
Expressivity (EPAQ)
Mastery (WOFO)
Work (WOFO)
Competitiveness (WOFO)

Appendix D - Chiefs of the Astronaut Office

Chief and Deputy Chief of the NASA astronaut Office (1962-2010)

1. Deke Slayton (September 1, 1962 - November 1963)

2. Alan Shepard (November 1963 - July 1969)

3. Tom Stafford (July 1969 - June 1971) (Stafford held the position while Shepard prepared for and flew Apollo 14)

4. Alan Shepard (June 1971 - August 1, 1974)

5. John Young (January 14, 1974 - April 15, 1987), Deputy was Paul J. Weitz. Acting Chief during STS-1 training was Alan Bean.[1]

6. Dan Brandenstein (April 27, 1987 - October 1992), Deputy was Steven Hawley.

7. Robert Gibson (December 8, 1992 - September 6, 1994), Deputy was Linda Godwin.

8. Robert Cabana (September 6, 1994 - October 1997), Deputy was Linda Godwin.

9. Kenneth Cockrell (October 1997 - October 1998)

10. Charles Precourt (October 1998 - November 2002), Deputy was Kent Rominger.

11. Kent Rominger (November 2002 - September 2006), Deputies were Andy Thomas and Peggy Whitson.

12. Steven W. Lindsey (September 2006 - October 2009), Deputies were Janet Kavandi and Sunita Williams (February 2008 to October 2009).

13. Peggy Whitson (October 2009–present) Deputy is Chris Ferguson.[3]

source: http://en.wikipedia.org/wiki/Chief_of_the_Astronaut_Office

Team Training for LongDuration Missions in Isolated and Confined Environments:

A Literature Review, Operational Assessment & Recommendations for Practice and Research

Raymond A. Noe, Ph.D.
Ali McConnell Dachner
Brian Saxton

Department of Management and Human Resources
Fisher College of Business
The Ohio State University

Prepared for NASA

September 2010

Please direct all correspondence to:

Raymond A. Noe
Robert and Anne Hoyt Designated Professor of Management
Department of Management and Human Resources
700 Fisher Hall
Fisher College of Business
The Ohio State University
2100 Neil Avenue
Columbus, OH 43210
614.292.3982
noe@cob.ohio-state.edu

Introduction

The Behavioral Health and Performance (BHP) element addresses human health risks in the NASA Human Research Program (HRP). BHP supports and conducts research to help characterize and mitigate risks for long-duration missions, and in some instances, current flight medical operations.

Although crew members and the ground crew currently receive training, additional training capabilities will be required for future exploration missions to Mars. These missions will have substantially different requirements for success than any previous NASA mission, so training will have to be revised accordingly. To ensure crew safety and accomplish mission work tasks, effective application of the skills and knowledge learned in training is critical. There is a need to understand recent developments in the team training literature as well as current team training strategies to help direct future training efforts in preparation for long-duration missions.

This report provides the results of a literature review on team training, operational assessment, evaluation, and recommendations for NASA's current team training strategies and future research that are relevant to the Team Risk (specifically focusing on monitoring task performance, psychosocial performance, and teamwork).

Literature Review

The purpose of the literature review was to identify research on current team training strategies including general models of training but also specific strategies for team training in isolated, confined, and extreme environments. Teams are defined as a distinguishable set of two or more individuals who interact dynamically, adaptively, and interdependently; who share common goals or purposes; and who have specific roles or functions to perform (Salas et al., 1992). There are several different types of teams (e.g., top management teams, task forces, surgical teams, shuttle crew teams, sports teams). Teamwork is defined by a set of interrelated knowledge, skills, and attitudes that facilitate coordinated, adaptive performance and support one's teammates, objectives, and mission (Alonso et. al., 2006; Baker, Gustafson, et al., 2003; Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Salas, Dickinson, Converse, & Tannenbaum, 1992; Salas, Bowers, & Cannon-Bowers, 1995). Teamwork depends upon each team member's ability to: 1) anticipate the needs of others; 2) adjust to each other's actions and to the changing environment; and 3) have a shared understanding of how a procedure should happen to identify when errors occur and how to correct for these errors. Marks, Mathieu, & Zaccaro (2001) identified three dimensions of teamwork behavior that have been empirically supported (e.g., Lepine, Piccolo, Jackson, Mathieu, and Saul, 2008). The dimensions include transition behaviors related to evaluating and/or planning to guide the accomplishment of a team goal or mission (mission analysis, goal specification, strategy formulation and planning), action behaviors or activities leading to goal accomplishment (monitoring, backup, and coordination behaviors), and interpersonal processes (conflict management, motivation, and confidence building, and how they affect management).

In team-based work environments, team members must have the knowledge, skills, and abilities (KSAs) that allow them to communicate and coordinate with other team members and perform complex tasks that require integration of team members' competencies (Delise, Gorman, Brooks, Rentsch, & Steele-Johnson, 2010). Team members also are expected to use their KSAs and perform tasks in stressful situations such as emergencies, time pressure, in distributed team environments, and facing information overload or deficiencies. As a result, team training is believed to be critical for effective team performance (Hollenbeck, DeRue, & Guzzo, 2004). Team training is a planned effort administered in a team environment to improve team performance (Goldstein & Ford, 2002; Klein et al; 2006; Noe, 2010). Team training is especiallyimportant for aviation, medicine, and space teams who all share the need for decision making based on incomplete or conflicting information, the need for coordination among professionals with different skills and ranks, and the likelihood that poor team performance will lead to serious consequences or death (Alonso, 2006; Cannon-Bowers & Salas, 1998). Team performance is defined as an emergent phenomenon resulting from a goal-directed process whereby members draw from their individual and shared resources to display task work processes, teamwork processes, and integrated team-level processes to generate products and provide services (Kozlowski & Klein, 2000; Salas, Stagl, Burke, & Goodwin, 2007). Effective team training is typically evaluated by determining the relationship between team training and one or more outcomes including cognitive, affective, process, and performance outcomes.

The literature review is organized as follows. First, an overview is provided of the process used to identify the articles and chapters included in this review. Next, the paper provides a general summary of team training and team effectiveness literature. This is followed by a discussion of team training research related to cross-training, training in isolated and confined environments, team mental models, and cross-cultural training.

Literature Review Process

The articles and book chapters included in this literature review are based on a search of electronic databases including journals from business, psychology, aviation, medicine, and engineering. The key words used to search these journals included team training, cross-cultural training, team mental models, isolated work, confined space, extreme environment, Antarctica training, aviation training, and environmental medicine. Also, additional articles were identified by examining the reference lists of these articles. As a result of this process, ninety-seven articles from forty-one journals were identified and included in this review. A complete list of journals included in the review is found in Appendix 1. An Excel spreadsheet and a Word document are available from the authors. The Excel spreadsheet includes a citation and abstract for each article and chapter reviewed. The Word document includes a detailed summary of each article including citation and a narrative describing the hypotheses or research questions addressed, study results, and implications for research and practice.

Overview of Research on Team Training Effectiveness

A series of meta-analyses strongly suggest that team training has a positive influence on team effectiveness. Delise, Gorman, Brooks, Rentsch, & Johnson (2010) conducted a meta-analysis of studies of the effectiveness of team training conducted between 1986 and 2007. They found that team training had a positive relationship with team effectiveness (as determined by *d* or effect size). Overall team training was found to be related to team outcomes (*d*=.85). Team training was positively related to affective outcomes (d=.80), cognitive outcomes (*d*=1.37), subjective task-based skill outcomes (*d*=.88), objective task-based skill outcomes (*d*=.76), and teamwork skill outcomes (*d*=.64). The differences in effect sizes were not significant, suggesting that team training did not have a significantly stronger relationship with any one type of effectiveness outcome. Team training also has been shown to be related to improvements in specific team processes and team skills. Based on their meta-analysis of team training including Crew Resource Management training, cross-training, guided team self-correction strategies, scenario and simulation-based training and team building, Klein et al. (2006) concluded that team adaptation and coordination training and Crew Resource Management training were the most effective in improving team performance and the performance of team behaviors, especially communication and coordination behaviors ($r=.629$). Team training interventions had a larger impact on team processes and performance than team member affective outcomes. Salas, DiazGranados, Klein, Burke, Stagl, Goodwin & Halpin (2008) found that team training was useful for improving cognitive, affective, process, and performance outcomes. Across all outcomes, team training interventions were more effective for team processes than for any other types of outcomes. Team training with a mixed training content (focus on teamwork and task work) was not found to be superior to those focusing either on teamwork or task work. The stability of team membership moderated the relationship between team training and team outcomes such that intact teams that underwent training improved the most on process and performance outcomes.

Team Training Generic Team Training and Team Training Methods

There have been a number of efforts to develop team training programs and recommendations regarding the skills, design, and delivery methods that are most effective. Ellis, Bell, Ployart, Hollenbeck, & Ilgen (2005) emphasize that the success of team training programs depends on conducting a thorough team training analysis, starting with a skills inventory to identify the competencies that are needed (Salas, Burke, & Cannon-Bowers, 2002). According to Ellis et al. (2005) competencies can be categorized into one of four groups depending on whether they are specific or general to a particular team and specific or general to a particular task. Past research has identified five categories of task- and team-generic competencies: (a) conflict resolution, (b) collaborative problem solving, (c) communication, (d) goal setting and performance management, and (e) planning and task coordination. *Planning and task coordination* refer to team members' capacity to effectively sequence and orchestrate activities, as well as manage procedural interdependencies among team members. *Collaborative problem solving* refers to team members' capacity to effectively use collective induction and deduction to resolve challenges and difficulties. *Communication* refers to team members' capacity to understand information exchange networks and to utilize these networks to enhance information sharing. Using students participating in the Distributed Dynamic Decision Making (DDD) simulation, Ellis et al. (2005) found that generic teamwork skills training significantly increased declarative knowledge within the team, and that trained teams demonstrated significantly greater proficiency than untrained teams in the areas of planning and task coordination, collaborative problem solving, and communication in a novel team and task environment. The DDD is a dynamic command and control simulation requiring team members to monitor activity in a geographic region and defend it against invasion from unfriendly air or ground tracks that enter the region. The training was not task- nor team-specific, and provided participants with no information regarding situations that they might encounter in the DDD simulation. The training also was conducted at the individual level. That is, team members were trained individually without any interaction with their soon-to-be teammates.

Stachowski, Kaplan, and Waller (2009) examined the relationships between characteristics of team interaction patterns and team effectiveness during crisis events. Crisis events were defined as “low-probability, high–impact events that are characterized by time pressure and ambiguity and that have significant consequences for an individual, team, and/or organization” (Yu, Sengul, & Lester, 2008, p. 252). They studied 14 intact nuclear power plant control room crews. The 14 crews were participating in a regularly scheduled training simulation. The simulation included simulated crisis events designed to portray realistic scenarios that are often based on events that occur at other plants. Effective teams are able to shed routinized, rigid interaction patterns and, as a result, are better able to adapt to emerging crisis situations. Results showed that higher performing crews exhibited fewer interaction patterns than did the less effective crews (“regular sets of verbalizations and non-verbal actions intended for collective action and coordination”). More effective crews engaged in less actor switching (two-way exchange of information), involved fewer team members in their patterns, and engaged in shorter, more concise interaction patterns that contained fewer behaviors than patterns of less effective crews. Superior crews exhibited fewer, shorter, less complex, and more flexible patterns of crisis response than did average performing crews. The results of this study highlight the limits of training teams to respond in a highly procedural manner or to adhere necessarily to an established pattern of interaction. The disadvantage of training that does emphasize adherence to specific procedures is that it may reduce trainee’s awareness of the need to deviate from these patterns and prevent them from acquiring the skills that would foster such deviation. Post hoc analysis of videos of crews in simulations suggested that the most effective crews used protocols as tools but did not allow them to guide their pattern of interaction. Stachowski et al. (2009) suggest that training should foster team interaction that is briefer and involves fewer actors and less back and forth communication. Training designed to teach teams to engage in briefer, more directive, and less inclusive interactions without sacrificing team knowledge would also seem useful.

A study by Katz-Navon, Naveh, and Stern (2009) suggests that it may be naïve to conclude that an active learning climate by itself is enough to ensure learning occurs that meets organizational objectives, especially in high-risk jobs and high reliability industries. Their study involved resident physicians in the health care industry and their medical treatment errors. In general, resident physicians have responsibility for the health and well-being of patients but, at the same time, are in the process of learning their profession. The authors proposed and investigated a multilevel model of how an active learning climate (a department-level phenomena) influences the number of errors (individual level), testing the moderating effect of safety priority and managerial safety practices (department level). They found that an active learning climate was associated with a greater number of errors. The interaction between an active learning climate and a priority of safety was significant, suggesting that a highly active learning climate with an immediate level of priority of safety was related to a low number of errors. They also found a significant interaction between active learning climate and managerial safety practices showing that the higher the active learning climate, the fewer the treatment errors when managerial safety practices were high. Their results suggest that different aspects or dimensions of a safety climate have a differential impact on error rates.

For flight crews on a long-duration mission in a confined environment, a needs assessment and task analysis are necessary for identifying the team skills that should be emphasized in training.

These skills likely include conflict resolution, collaborative problem solving, communication, goal setting and performance management, and planning and task coordination. Team training for long-duration missions should ensure that flight crews and flight controllers are able to shed routinized patterns and become more flexible to deal with crisis events.

Simulations, software, virtual worlds. Several articles discuss the use of simulations, virtual worlds, and software for aviation and medical team training (Hamman, 2004; Heinrichs, Youngblood, Harter, & Dev, 2008; Krauss and Gramopadhye, 1999; Lerner, Magrane, & Friedman, 2009; Shapiro, 2004). For example, Hamman (2004) discusses the implications of aviation team training for medical team training. Two primary types of training are discussed: the Advanced Qualification Program (AQP) and the simulation scenario design process. Hamman (2004) emphasizes that simulator design must be interdisciplinary in focus, requiring real communication. Discrete events should be identified and tested, and specific skills should be identified for each event. Team training skills must be identified by task analysis, have identified Topic Proficiency Objectives (TPO), Supporting Proficiency Objectives (SPO) skills, and behavioral markers of performance. From the first day of training, team skills should be integrated into the curriculum lesson plans and supported by curriculum design. Team training skills must share equal importance with the technical skill requirements. Hamman (2004) recommends that the curriculum must be designed to support cross-cultural training and must integrate a carefully designed simulation that is based on scientific models of team training generated from performance data from the environment. The team training elements must be integrated into the event set design with defined criteria for successful outcomes.

Kraus and Gramopadhye (1999) examined the role of team training and the use of advanced technology in the aircraft maintenance environment. As part of the research, computer-based team training software (Aircraft Maintenance Team Training (AMTT) software) was developed. In this study, usefulness of AMTT was tested against a traditional classroom method of instruction in terms of team knowledge, acquisition, and usability issues. They found that there were no significant differences in user satisfaction between instructor-based training and computer-based training. Subjects with low levels of computer literacy were able to interact and use the AMTT software after minimal instructions on basic computer operations. Computer-based training was as effective in delivering team training instruction as instructor-based training.

The use of simulations, virtual worlds, and software for team training is promising. These methods will be especially important on long-duration missions because crew members will be responsible for "learning as they go" (on-board learning) to refresh previously trained skills or to acquire new skills to deal with unexpected crisis or events. Regardless of when, how, and where team training occurs, it should receive at least similar level of emphasis and importance as technical skills, and, to the extent possible, be integrated into operational training

TeamSTEPPS. The Department of Defense and the Agency for Healthcare Research and Quality developed TeamSTEPPS, an evidence-based curriculum. TeamSTEPPS evolved from research in high-risk fields such as aviation and aeronautics, nuclear power, and the military, where poor performance can lead to serious consequences (TeamSTEPPS, 2005). TeamSTEPPS focuses on the core principles of teamwork identified by researchers like Mumford, Zaccaro, Harding, Jacobs, and Fleishman (2000) and Kraiger, Ford, and Salas (1993) by teaching specific tools and strategies that can be used to improve teamwork performance in the military medical environment. Core skills of TeamSTEPPS include leadership, situation monitoring, mutual support and communication. Leadership refers to the ability to coordinate the activities of team members by ensuring that team actions are understood, changes in information are shared, and team members have the necessary resources. Situation monitoring is the process of actively scanning and assessing situational elements to gain information, understanding, or mutual awareness to support team functioning. Mutual support refers to the ability to anticipate and support other team member's needs through accurate knowledge about their responsibilities and workload. Communication is the process through which team members clearly and accurately exchange information. TeamSTEPPS is being integrated into obstetrical emergency training (see Daniel & Simpson, 2009).

Fox, Johnson, Gagliano, Passarello, Moore, Resurreccion and Reed (2006) used TeamSTEPPS to train senior surgical residents on teamwork skills. Generally speaking, the residents do not typically receive training in leadership or teamwork skills. Residents attended one 4-hour training session followed by attending a trauma conference.

Following training, two surgical grand rounds were dedicated to reinforcing team training skills. Comparison of pre- and post-training surveys of the residents showed that residents perceived improvement in the team's ability to measure performance. Also, they felt that team roles were better defined, the team worked well together, communicated more effectively, and perceived an improvement in the team's ability to resolve conflict.

Research on TeamSTEPPS suggests that team training in leadership, situational monitoring, mutual support, and communications can help flight crews and flight controllers more clearly understand their roles and enhance communications, coordination, and conflict resolution.

Team cognition and team effectiveness. Emerging research suggests that team cognition is important for team performance and team effectiveness and may be developed through team training. Team cognition is an emergent state that refers to the manner in which knowledge important to team functioning is mentally organized, represented, and distributed within the team allowing team members to anticipate and execute actions (Kozlowski & Ilgen, 2006). Two important cognitive constructs have been identified as being important for teams. First, team mental models or shared mental models are a "team members' shared, organized understanding and mental representation of knowledge about key elements of the team's relevant environment" (Mohammed & Dumville, 2001, p. 90). Shared or team mental models can be considered mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states. Teams whose members share models of both task work and teamwork can better anticipate the needs and actions of other team members resulting in better team performance. Teams with a well-developed team mental model should have a common view of events and incidents, what they are likely to lead to or cause, and why they are occurring. Team mental models are a mechanism through which the team members can coordinate actions and adapt behaviors leading to improved decision making and performance. Team mental models are a property of teams that emerge as a function of team member characteristics, the context, processes, and outcomes. Second, transactive memory refers to knowledge that is distributed among team members. Team mental model similarity refers to the extent to which team members' mental models are shared, consistent, or converge among team members.

Cannon-Bowers, Salas, & Converse (1993) proposed that a team is most likely to be effective if team members share four non-independent mental models. The equipment model captures team members' shared understanding of the technology and equipment with which they carry out their team tasks (task work). The task model captures team members' perceptions and understanding of team procedures, strategies, task contingencies, and environmental conditions (task work). The team interaction model reflects team members' understanding of their responsibilities, norms, and interaction patterns (teamwork). The team model summarizes team members' understanding of each others' knowledge, skills, attitudes, strengths, and weaknesses (teamwork). Each mental model may be influential in predicting team performance. Findings suggest that teams whose members structure and organize their team-related knowledge in a similar fashion are likely to find it relatively easy to coordinate their activities. They are likely to agree upon team priorities and strategies, yielding efficient task performance. In practice, researchers have tended to collapse team mental model content into teamwork categories (interpersonal interaction requirements and skills of other team members) and task work categories (work goals and performance requirements) (see Mohammed, Ferzandi, & Hamilton, 2010) .

Recent empirical evidence suggests that mental model similarity improves team coordination processes, which in turn enhance team performance (Marks et al., 2000; Mathieu et al., 2000). Unlike Mathieu et al. (2000), Lim and Klein (2006) found a direct relationship between team mental model similarity and team performance. This may reflect the high stress and intense time pressure context in which the teams that they studied were trained to operate. Under such circumstances, there is very little time for explicit coordination and communication. To succeed in their tasks (e.g., reacting to an enemy's ambush), team members must have a shared understanding of the emerging situation and the collective action required. It is precisely in this type of context that shared mental models have been hypothesized to be most predictive of team performance. Lim and Klein's (2006) results also suggest that team mental model accuracy is instrumental for team performance. Teams whose average mental models were most similar to experts' mental models performed better than did teams whose average mental models were less similar to experts' mental models. Team mental model convergence has been shown to be related to team processes (backup behavior quantity and quality, coordination, communication), emergent states (team collective efficacy, norms) and effectiveness (performance, viability, team member growth, strategy implementation) (see Mohammed et al.'s (2010) 15-year review of research on the team mental model construct).

Research has examined how team mental models evolve over time. McComb (2007) suggests that team mental model convergence proceeds through three phases: orientation (becoming familiar with the team situation), differentiation (creating unique views of the situation), and integration (allowing team member perspectives to develop into a collective focus). Langan-Fox (2003) suggests a skill acquisition framework for the development of team mental models involving orientation/negotiation (acquiring facts about the task and team), refinement/learning (constructing skills through processes and interaction), and high performance (expert team mental models).

DeChurch and Mesmer-Magnus's (2010) meta-analysis examined three questions: First, how important is cognition to team performance? Second, what aspects of cognition are most important for team processes and performance? Finally, which types of teams benefit from team cognition? The study examined both the broad relationships among team cognition, behavior, motivation, and performance outcomes as well as potential moderators of these relationships. They found a positive relationship between cognitive and behavioral processes overall (.43), as well as between cognition and both transition and action processes (r=.43 and .29, respectively). They also found a positive relationship between cognition and overall motivational states (.37) and, more specifically, between cognition and cohesion (.40). Their results also suggest a positive relationship between team cognition and team support (.38) and show that cognitive cognition makes a unique contribution beyond team cohesion and team behavioral processes in understanding team performance. Compositional cognition was more predictive of process for action and decision-based teams and most predictive of performance in project and decision making teams. These results emphasize that team cognition is an important team property, and training is needed to shape the collective cognition needed for effective teamwork.

It is fairly clear from the large number of studies of team mental models that developing shared mental models for flight crews and controllers is critical for long-duration missions in which crisis may leave little time for explicit coordination and communications.

Training to develop team cognition: Cross-training and cross-understanding. Training is the primary mechanism for enhancing team mental model development. Various types of team training, including self-correction, team-interaction training, computer-based, and cross-training, have been found to increase team mental model similarity and accuracy.

How might team cognition be developed to enhance team performance? Pearsall, Ellis, & Bell (2010) found that role identification behaviors occurring in the initial stages of team development were positively related to team mental models and transactive memory development. Role identification behaviors refer to purposeful interpersonal interactions directed toward understanding their teammates' roles and capabilities. Team members share information regarding their specialized knowledge and skills and abilities with the rest of the team. The degree to which team members engaged in role identification behaviors predicted the development of team interaction mental models. These cognitions mediated the effects of role identification behaviors on team performance during team compilation, as the coordination gained through the exchange of role-based behavior led to more effective and efficient teamwork. Similarly, Huber and Lewis (2010) emphasize cross-understanding, the extent to which the group's members posses accurate perceptions of the mental models of other members, as important for team effectiveness. This differs from team mental model studies that have focused on the extent to which team members actually do share teamwork and task work models rather than on their perceptions of sharing. Huber and Lewis (2010) discuss how different levels and distributions of cross-understanding affect group performance and learning. They differentiate cross-understanding from transactive memory system, emphasizing that it does not depend on, nor does it necessarily lead to, a division of cognitive labor. Huber and Lewis (2010) suggest that the challenge is determining how to staff teams to obtain both the likelihood of a high-quality group product that would occur from members having diverse mental models and smoothly coordinated processes that would follow from cross-understanding.

Cross-training is a type of team training in which team members rotate positions to develop an understanding of the basic knowledge necessary to successfully perform the tasks and duties of other team members. Research shows that cross-training appears to have a positive influence on the development of shared mental models but is less effective than other types of team training in improving team effectiveness. Marks, Sabella, Burke & Zaccaro (2002) studied the impact of the cross-training of action teams on team effectiveness. An action team is any team in which expertise, information, and tasks are distributed across specialized individuals, where team effectiveness depends on rapid, complex, and coordinated task behavior, and the ability to dynamically adapt to the shifting demands of the situation (Kozlowski et al., 1996). Action teams contain more specialized skill sets, rely more heavily on coordination, perform in less familiar and more challenging environments, and may be more temporary than traditional teams.

Marks et al (2002) proposed that cross-training influenced the development of shared mental models among team members, which in turn facilitated development of coordination and backup procedures and team performance. Marks et al (2002) identified three different types of cross-training: positional clarification, positional modeling, and positional rotation. The least in-depth form of cross-training**,** *positional clarification*, involves verbally presenting team members with information about their teammates' jobs through lecture or discussion methods. *Positional modeling*, entails both verbal discussion and observation of team members' roles. *Positional rotation* provides a hands-on approach to learning interpositional information by giving members experience carrying out teammates' duties through active participation in each member's role. Individuals are provided with training and first-hand experience of their team members' roles. This type of training parallels the concept of job rotation. They conducted two studies using student samples engaged in simulations. They found that cross-training significantly influenced development of team mental models, team mental models accounted in significant variance in team backup behavior and performance, and positional modeling and positional rotation were superior to positional clarification in terms of teammates having a greater understanding of each other's responsibilities. They also found that teams receiving positional rotation were more comfortable switching roles than teams receiving positional modeling and clarification, and teams receiving positional modeling and positional rotation developed mental models with a higher percentage of shared team interaction knowledge than teams who received positional clarification. The relationship between shared mental models and team performance was completely mediated by team coordination.

Using meta-analysis, Salas, Nichols, & Driskel (2007) compared the effectiveness of cross-training, team coordination and adaptation training, and guided team self-correction. Guided team self-correction refers to team training in which team members learn to diagnose the team's problems and to develop effective solutions. Guided team self-correction training is assumed to help develop correct expectations (i.e., shared mental models) among team members, therefore contributing to more effective performance. Smith-Jentsch , Cannon-Bowers, Tannenbaum and Salas (2008) found that guided team self-correction organized around an expert model of team work resultsw in more accurate but not more similar mental models of team work. Team coordination and adaptation training refer to team training in which team members are asked to alter their coordination strategy and reduce the amount of communication necessary for successful task performance. Salas et al. (2007) found that cross-training (r=-.09) was not as effective as self-correction training (r=.45) and team coordination and adaptation training (r=.61).

Team training has a positive influence on the development of shared mental models and team performance. Analogues, simulations, and other instructional methods used to train teams for long-duration missions should incorporate experiences that facilitate the development of guided team self-correction and the ability to alter their coordination strategy.

Multi-Team Systems (MTS)

Multi-team systems (MTS) are two or more teams that interface directly and interdependently in response to environmental contingencies to accomplish goals (Mathieu, Marks, & Zaccaro, 2001). For example, mission accomplishment depends on both the effective coordination and communication amongst flight controllers on the ground and the flight crew on the space vehicle as well as between these two teams. Marks, DeChurch, Mathieu, Panzer, & Alonso (2005) studied MTS composed of a leader team and two operational teams. Marks et al (2005) investigated team work processes that occurred during two phases of team performance (action phase and transition phase). Transition processes include planning, mission analysis, and goal specification. Action processes include monitoring progress toward goals, system monitoring, team monitoring, and backup behavior and coordination. Cross-team processes predicted MTS performance beyond that accounted for within team processes. Cross-team action processes were more important for MTS effectiveness when there were high cross-team interdependence demands. Positive transition processes related significantly to MTS performance both directly and mediated by MTS action processes. In a longitudinal study, Hoegl, Weinkauf, and Gemuenden (2004) studied multi-team research and development projects. They found that collaborative processes between teams during the project predicted later team performance. Inter-team coordination was especially important for teams that had technical interfaces with other teams. Collaboration both within and between teams in the early project phases effected subsequent performance. Based on the study results, Hoegl et al (2004) suggest that managing inter-team coordination, project commitment, and teamwork quality early on in the project helped to detect and counteract problems before project-controlling instruments are able to detect deviances.

DeChurch and Marks (2006) studied leadership in multi-team systems. Leaders were trained in two forms of process facilitation: strategy development and coordination, and the leaders' interactions with the two teams were examined. Strategy training had a stronger effect on explicit coordination. Coordination training had a stronger effect than strategy training on implicit coordination between teams. MTS leaders trained in inter-team planning and coordination skills were able to align and integrate efforts across teams, resulting in superior MTS performance.

The success of a long-duration space mission is dependent on the collaboration and coordination not only within flight crews, flight controllers, and other teams on the ground, but between these teams that make up a larger multi-team system. The small but developing research on multi-team systems suggests that team training for long-duration missions should emphasize both within and inter-team coordination and collaboration.

Team Training in Isolated and Confined Environments

Several researchers addressed the psychological and teamwork issues that individuals face in polar expeditions and polar analogue training. Palinkas and Suedfeld (2008) describe the psychological effects of polar expeditions, which included sleep deprivation, affectual changes, and interpersonal conflict. Gouriant, Apel, and Delbart (2010) provide a brief review of a space-type training mission that occurred in a polar outpost in 2007.

Ball and Evans (2001) emphasize that astronauts on long-duration space missions will confront a range of intra- and interpersonal challenges, the nature of which cannot be accurately determined at present. Therefore, substantive features of training must be based on continuously accumulating experiences in actual space flight environments and analog settings. High-fidelity training experiences should be developed based on specially designed algorithm software packaging technologies that accurately model space flight experiences and outcomes of flight crew actions. Naturalistic studies on the efficacies of specific training procedures must follow in both simulated and actual space mission settings. Additionally, they emphasize that personalized individual training approaches must also incorporate and evaluate countermeasures based on procedures for evaluation of cognitive and behavioral functioning that are adaptable for computerized administration as self-assessment and supportive intervention procedures (work cited by Wolpe, 1958; Beck and Emery, 1985; Power et al., 1990; Beck, 1993; Barlow, 1996; Cautela and Ishaq, 1996; Rosen and Schulkin, 1998; Lazarus, 2000). These programs were designed within a stress management context and were effective when combined with a range of interventions including biofeedback, relaxation techniques, systematic desensitization, and pharmacological treatments. Ball and Evans (2001) suggest that empirical studies of both individual and team behaviors in simulated flight exercises, conflict resolution strategies, and cockpit resource management programs can help increase understanding of how behavioral patterns influence performance effectiveness and guide decisions about group composition and training.

In addition, they provide several recommendations for training, including an integrated approach that includes ground-based monitoring and support groups specifically selected to participate in such operations. Firstly, they suggest that NASA behavioral health personnel should be directly involved in crew selection and in training crewmembers and ground control personnel in crisis intervention and problems with interpersonal functioning. They also emphasize that appropriate assessment tools and countermeasure development are required to address emergencies and technical assistance requirements under conditions that involve multinational crews and the complexities related to cultural and language differences, as well as under conditions that involve crews of mixed sexes and with command structure constraints (work cited by Kelley and Kanas, 1992; Holland, Looper, and Marcondes-North, 1993). Within this context, it is not enough to have the leader be the buffer, because the leader could be addressing specific problems or could be too involved in a task-oriented emergency. Finally, crew resource management for long-duration missions also requires consideration of technical as well as nontechnical skills (e.g., corporate citizenship, interpersonal skills, and compatibility). Individual differences in personality functioning become important when the job requires corporate citizenship or the use of "people skills" (Borman et al., 1997; Hogan et al., 1998; Mount et al., 1998; Salgado, 1998). They recommend use of the distributed interactive simulation methodology. Distributed interactive simulation environments are based on multi-person computer-generated workstation networks that represent operational elements consisting of both individuals and functional groups. Such techniques involving real people can be used for selection and training under conditions of simulated mission operations in a realistic environment. Participants communicate via electronic channels to exchange information, discuss work requirements, and evaluate data for decision making; exchange the outcomes of specific actions; and evaluate mission-oriented scenarios. Space mission simulations also permit inquiries of Earth-based mission control for information or instructions, or both. Groups of individuals are trained to interact within the simulation environment for the purpose of engaging with assigned crewmembers and Earth-based mission control. Distributed interactive simulation methodologies with performance tasks requiring repeated exchange of information among participants and between groups provide an automated means for the systematic monitoring and analysis of the effects of experimental variations on psychosocial interactions, decision making, and both individual and group performance effectiveness. The operational performance measures evaluated include pattern analysis, task completion, and timing parameters.

Similarly, Kanas, Sandal, Boyd, Gushin, Manzey, North, Leon, Suedfeld, Bishop, Fiedler, Inoue, Johannes, Kealey, Kraft, Matsuzaki, Musson, Palinkas, Salnitskiy, Sipes, Stuster, and Wang (2009) identify specific training issues that need to be considered in isolated confined environments during long-duration space missions. These include participation of all space agencies, training in self-care and self-management, teamwork, and group living, language training, sensitivity training, leadership and followership, and cross-cultural training. Kanas (2004) also emphasizes the need to sensitize astronauts to interpersonal issues, to be trained to monitor themselves for interpersonal problems, and to develop problem-solving exercises to be used in training that involve both astronauts and ground control.

Dion (2004) emphasizes that crew selection, training, and backup plans are critical for the success of long-term space missions. He also emphasizes that team cohesion is very important to success and should be reinforced in training. Team cohesion needs to be monitored continuously, the crew needs to be trained in cohesion-building skills like team interventions, and training needs to build team identification rather than subgroup identification to avoid problems of in- and out-group dynamics.

Orsanu (2005) cautions that although training is important, many events that space crews face will be unique and unpredictable, emphasizing the importance of good decision making in the face of uncertainty. Poor decisions are likely to occur when relevant knowledge of the problem is not available, available information is of poor quality, accurate information is available but it is difficult to interpret, and inaccurate projection of decision consequences has taken place (which means that mental models of the situation are important). She also emphasizes that training is important because it increases self-efficacy and increases metacognitive strategies like planning ahead for difficulty later. Training should explicitly address what decision making will be encountered under stress. Lastly, she summarizes the work of Lipshitz and Strauss (2001), who in their studies of military decision makers, found five principal strategies for coping with uncertainty (RAWFS): 1) Reducing uncertainty by collecting additional information; 2) Assumption-based reasoning (filling gaps in knowledge by making assumptions that go beyond directly available data); 3) Weighing the evidence of at least two competing hypotheses; 4) Forestalling (developing an appropriate response or response capabilities to anticipate undesirable contingencies); and 5) Suppressing uncertainty (e.g., by ignoring it or by relying on unwarranted rationalization).

Flight crews and controllers involved in long-duration missions cannot be trained before the mission for every experience they might encounter. As a result, flight crews and controllers need to be provided with a strategy for approaching crisis and uncertainty. In preparation for long-duration missions, flight crews need to be evaluated on self-care, communications, teamwork, contributions to team cohesion, decision making, crisis management, and cultural agility.

Crew Management Training, Crew Resource Management Training (CRM) and Space Flight Resource Management Training (SFRM). O'Connor, Campbell, Newon, Melton, Salas, and Wilson (2008) conducted a meta-analysis of studies of CRM effectiveness. Their results supported the effectiveness of CRM. CRM had a positive relationship with trainee reactions, attitude change, and knowledge acquisition. It is important to note that these results should be interpreted with caution because only a small number of studies of CRM effectiveness (n=16 out of 74 total studies) had sufficient data (e.g., correlations, effect sizes) for inclusion in the meta-analysis.

One important issue in CRM and SFRM is to clearly identify the behaviors that should be observed and which behaviors are appropriate for specific situations. Flin and Martin (2001) note that CRM has been required for some time for pilots, but there is little evidence as to its effectiveness because the incidents of accidents are too low to have sufficient variance to study. They suggest that establishing behavioral markers and having instructors or trainers provide ratings of crew members on these markers is one way to evaluate CRM effectiveness. The behavioral markers should be based on the cognitive and interpersonal skills emphasized in CRM. The cognitive dimension of CRM training includes situational awareness, workload management, planning, and decision making which make up most problems according to instructors. Interpersonal dimensions include crew coordination, communications, leadership, and group climate. During simulations or actual flight, it is important that event sets are clearly specified to ensure that instructors agree and are aware of the behaviors the crew should be exhibiting to demonstrate competence in a specific CRM dimension. Also, it is important that instructors be asked to identify actions that indicate that a decision has been made, not the actual decisions. Because evaluation of the behavioral markers is based on subjective assessment of trainers/instructors, Flin and Martin (2001) emphasize that rater training is critical. They also emphasize that the best way to evaluate CRM training is to observe the crew during simulation or flight. In their study of MC-130P aircrews, Nullmeyer & Spiker (2003) found that seventy-five percent of the observed variability in mission performance ratings was accounted for by ratings of CRM skills. The rated skills included functional allocation, tactics employment, situational awareness, time management, and command-and-control communications. Behaviors that were related to high ratings included giving greater consideration of the "big picture", viewing the crew as only part of the larger team and mission, raising extensive "what-if" questions about main mission events (including input from the entire crew), accepting the need to change the plan based on the evolving mission and changing situation, including explicit alternatives within permission briefings, and responding well to their own errors or changing conditions. Time management also was highly correlated with mission performance. Exceptional crews were aware of time and their use of time throughout mission planning and execution. Also, in the most effective crews, individual crew member's duties were overtly and explicitly designated based on crew member strengths rather than position. Other important process behaviors that did not fit into the CRM categories used in the study included mission focus, development and use of aggressive plans, and emergence of a clear, single leader who worked to integrate the crew together in all aspects of the mission.

Space Flight Resource Management training (SFRM) was originally modeled after the airline industries' and military's Crew Resource Management (CRM). It was designed to address the team skills required for crew members and flight controllers during time-critical scenarios found throughout a mission. SFRM training emphasizes eight interrelated team skills: Communication, Cross-Culture, Team Work, Decision

Making, Team Care, Leadership/Followership, Conflict Management, and Situational Awareness. SFRM also provides a technique to deal with momentary loss of situational awareness: Stop, Think, Act, Review (STAR). Before reacting to an event or beginning a task, the individual should "Stop" and take the time available to focus on what he or she is about to do. Next, they should "Think" about the situation at hand. What are the defining factors and critical circumstances of the situation? How is the situation similar and different from previous situations they have experienced? Once the individual has gained a clear awareness of the task or situation, he/she then should develop options including risks, consequences, worse-case scenarios, and contingency plans for each option. After deciding on an option, the individual must "Act" on the option using error-prevention techniques. During and at the completion of each step of the selected course of action, the individual is expected to "Review" the process and outcome. If the option does not go according to plan (or starts to show signs of deviating from the expected plan), then the individual starts the STAR process over again. SFRM helps teams increase situational awareness, learn to work together as a team, and check and back up other team members. It helps the crew to know how to handle situations where the only available resources are those on the spacecraft.

SFRM skills are important for flight crews and controllers involved in long-duration missions. Analysis is needed to determine which current SFRM skill or new skills are needed for effective performance on long-duration missions. It will be difficult for flight crews on long-duration missions to interact with the ground, and the types of events and crisis they will face may differ significantly from those experienced on Shuttle and the International Space Station (ISS). SFRM skills are especially important for increasing flight crews and controller's situational awareness and improving problem-solving skills needed to deal with emerging situations using the resources available on the spacecraft.

Cross-Cultural Training

There is a small but growing body of research and awareness of the importance of cultural differences in space missions and analog environments. Survey research and anecdotal evidence suggests that cultural differences between crew members can impede the effectiveness of space missions, especially when these missions involve astronauts from different international agencies (e.g., Kraft, Lyons, & Binder, 2003). Helmreich (2000) found that Hofstede's dimensions of individualism-collectivism and power distance were important determinants of error rates in aviation environments. Kealey (2004) presents key research findings about intercultural effectiveness and discusses its relevance for space missions, highlighting some of the issues that should be addressed to help minimize problems related to this intercultural effectiveness, and providing suggested action steps needed to address issues associated with multicultural functioning. Kealey (2004) defines intercultural effectiveness as "the ability to live contentedly and work successfully in another culture" (C58). He emphasizes that the success of multicultural crews may be more influenced by their interpersonal skills than their technical skills. Some of the issues identified and action steps provided by Kealey (2004) include:

- It appears that most people rate themselves as interculturally effective, even when their fellows and supervisors do not agree. This may explain why most people, across many analogue settings, are satisfied with their assignments, even if they do not possess effective multicultural skills.

- Individuals tend to interact with fellow crew members from their own culture; countermeasures for this “ingroup/outgroup” effect should be considered during training.
- Hardship tends to bond participants, which should help during long-term missions. Monotony may counteract this effect.
- Mission control is often viewed as outsiders making unrealistic demands; this needs to be addressed ahead of time as miscommunication with mission control can be dangerous.
- There has been some research on identifying the kinds of skills that facilitate intercultural success. However, insufficient attention has been given to contextual factors (what is “the right stuff” changes depending on the situation).
- Intercultural training, like other kinds of training, is increasingly focused on competencies; this approach can likely be adapted to training for space missions

Ritsher (2005) identifies the cultural differences between Russian and American space station crew members and provides training recommendations to increase crew member awareness of these issues. Some of the cultural differences highlighted by Ritsher (2005) include personality differences (e.g., extraversion and openness to experience are generally higher among Americans, Russians are higher in expressivity). In addition, Americans are more reliant on roles, as opposed to Russians who generally depend on personal relationships, Americans are used to more personal space, have different personal hygiene habits, and gender norms differ between Americans and Russians.

Ritsher (2005) recommends that specific training be designed to address these issues. Team-building exercises and other team training activities should deal explicitly with the cultural issues that the team configuration will face. The crew should be led to think about cultural differences before and develop strategies for dealing with them. Crew should be trained to act as a “psychological health officer” to evaluate the extent to which cultural differences are creating issues that are inhibiting crew effectiveness.

In a study of the European Space Agency using surveys measuring cultural factors and their effects, Sandal and Manzey (2009) found that national cultures do significantly impact ability to accomplish the mission, and cooperation within and between space agencies is important. These results suggest that despite the European Space Agency having its own organizational culture, national cultures still have a strong influence on cooperation. Also, Tomi, Kealy, Lange, Stefanowska, and Doyle (in press) found that mistrust between organizations was a major issue preventing cooperation, along with the usual miscommunications and differences in work style. Clement and Ritsher (2005) investigated cultural effects in mission control and found that they affected performance; the researchers suggested that strong communication efforts and robust relationships were important to overcome cultural differences. Based on interviews with flight controllers for the International Space Station (ISS), Clement, Boyd, Kanas, and Saylor (2007) found that Russian and American controllers have different approaches to documentation, planning, and problem solving. Both junior and senior flight controllers reported that it is necessary to be aware of cultural differences, try to accommodate differences, and look for clues that partners are operating under different assumptions.

It is important to note that there is a voluminous body of cross-cultural research in the management literature, some of which may be applicable for understanding cultural differences between crew members and training crew members (for example, see the review by Gelfand, Erez, & Aycan, 2007). This literature includes studies addressing different cross-cultural training methods, repatriation, and the predictors of success in cross-cultural assignments. For example, consider the recent work by Brandl and Neyer (2009) on virtual teams. Global virtual teams are culturally diverse, involve two or more nations, work across temporal and physical distance, are interdependent, and rely on technology-mediated communication (Baba, Gluesing, Ratner, & Wagner, 2004). Global virtual teams are challenged to overcome anxiety and uncertainty that influence the effectiveness of their communication (Gudykunst, 1995). Anxiety and uncertainty management (AUM) theory proposes that anxiety and uncertainty are central elements influencing the effectiveness of intergroup communication (Gudykunst 1995, 1998, 2005).

In cross-cultural interactions, according to AUM theory, the ability to manage uncertainty and anxiety are central elements of effective communication (Gudykunst, 1998). Communication becomes more difficult if uncertainty and anxiety are too high (Gudykunst, 1998). High levels of uncertainty and anxiety in cross-cultural interactions reduce one’s ability to predict and interpret the behavior of others (Brandl & Neyer, 2009). Effective communication also suffers when uncertainty and anxiety are too low. Low levels of anxiety and uncertainty result in overconfidence and decreased motivation to communicate, gain new knowledge, and accurately interpret cultural differences. Effective communication and establishing trust are especially difficult in technology-mediated interactions in global virtual teams (Jarvenpaa & Leidner, 1999). There are more misunderstandings among global virtual teams because they are characterized by a high degree of uncertainty and anxiety that needs to be overcome. To communicate effectively in cross-cultural settings, people must be open to new information, aware of alternative perspectives, and able to adjust quickly to unknown situations so they can make more accurate predictions about the behavior of others (Gudykunst, 1998).

Newcomers need to change their mental models as a requisite for effective communication in cross-cultural interactions (Gudykunst, 1998). To perform effectively in a context of uncertainty, global virtual team members must learn how to interpret the other cultures' vernacular and how to express themselves in the other cultures' vernacular to achieve goals. Additionally, team members must become familiar with the "letting go and taking on" (Osland, 2000, p. 235) strategy toward the perceptions of the other culture (Brandl & Neyer, 2009).Thus, the effectiveness of cross-cultural team communication and interactions depends on the reorganization of mental maps, adaptation of behavior to the intercultural situation, and being sensitive to the specific set of behaviors that is appropriate in the setting (Chen & Starosta, 1996). Social interaction can be supported by certain types of cross-cultural training.

Brandl and Neyer (2009) suggest that cultural awareness training will result in more effective cross-cultural communication among teams than traditional cultural orientation programs. Cultural orientation programs attempt to reduce uncertainty in unknown situations by educating newcomers about country-specific knowledge that basically equates to teaching cultural stereotypes. Brandl and Neyer (2009) argue that "country-specific knowledge is not a substitute for in-depth knowledge of interpersonal interactions" (346). Additionally, these authors also suggest that "ready-made concepts by themselves are not sufficient to capture the other team members' cultural pattern" (347). The proposed shortcomings of cultural orientation training are not an issue in cultural awareness training which, "seeks to enhance the team members' capabilities to adjust to unknown situations" (Brandl & Neyer, 2009, 347).

Cultural awareness training encourages participants to master unknown situations by seeking information to enhance their awareness of alternative perspectives (Brandl & Neyer, 2009; Gudykunst, 1998). The objective of cultural awareness training is to help newcomers better deal with unfamiliar situations when working together with people from different cultures by developing newcomers' openness to new information and their awareness of alternative perspectives (Gudykunst, 1998). Brandl and Neyer (2009) suggest that cultural awareness training facilitates the adjustment process to new situations in three ways (Glanz et al., 2001; Kohonen, 2004, 2005):

> First, team members become aware that uncertainty inevitably arises during their participation in global virtual teams. Second, as in cultural awareness training, when team members experience how to achieve solutions and activate supportive resources, they are more willing to explore unknown situations. Third, the complexity of team members' mental maps improved in this form of training enhances their ability to link schemata to contexts (348).

Earley and Peterson (2004) suggest that most cross cultural training has focused on country specific knowledge. They argue that the general approach to cross-cultural training suffers from several weaknesses. It assumes that everyone needs to know the same thing; it assumes a similar level of interaction at a common location. It also tends to focus on cognitive skills, giving less emphasis to the metacognitive skills needed to be adaptable. Finally, they criticize cross-cultural training emphasis on analogical learning because most individuals cannot transfer that kind of learning to new settings. They propose an alternative approach focusing on cultural intelligence. Cultural intelligence (CQ), as they conceive it, focuses around two primary issues: metacognition, motivation (in the face of failures), and behavior (mimicry and behavioral repertoire). For metacognition, training should focus on how to learn from experience and how to deal with new knowledge. Goal-setting training will be helpful to build motivation (e.g., set small goals and build toward larger goals). Role modeling and self-presentation training are good for developing appropriate behavior.

Cultural agility is the ability to quickly and comfortably work in different countries and with individuals from diverse cultures (Caligiuri, 2010). Cultural agility is important for flight crew members to develop credibility and communicate and work together effectively. Caligiuri (2010) suggests that individuals need three concurrent orientations to operate across countries and in multicultural settings. These orientations include cultural adaptation, cultural minimalism, and cultural integration.

Cultural agility is not cultural adaptation, however there are times when adaptation is critical. Cultural adaptation is an individual orientation to be sensitive and strive to adapt to the nuances of cultural differences, often leveraged in situations requiring individuals to behave in the most culturally appropriate ways to be successful. Cultural agility does not mean we should pretend cultural differences are nonexistent, however there are times when higher order demands will supersede cultural expectations. Cultural minimalism is an individual orientation to reduce the perceived influence of cultural differences either in one's own behavior or in the behavior of others. Cultural minimalism is a highly functional cultural orientation in situations where there are important strategic reasons to override or play down cultural differences. Cultural agility is not merely merging multiple cultures to create a new set of behavioral norms, but there are times when cultural integration is most important. Cultural integration is an orientation to understand the cultural differences of each person in a multicultural or cross-cultural context, but strive to create something that is a combination of many cultural perspectives.

Culturally agile individuals are able to operate with each of the three cultural orientations, depending on the situational demands. They will leverage the behaviors of a cultural minimalist when the situation demands that their behaviors supersede the local context. They will adapt their behaviors when the situation demands attention to the local context. They also will be able to create a new behavioral set taking elements from multiple cultural contexts. Cultural agility is gained over time as an individual builds a repertoire of appropriate responses and becomes more fluent in reading and assessing a given cultural context. Most of this learning is experiential as individuals interact with peers from other cultures and learn to test their assumptions and the limits of their personal knowledge.

It is highly likely that a long-duration mission will include a multinational multicultural flight crew and flight controllers. For flight crew and flight controllers to effectively communicate (verbally and non-verbally), live together, work together, and understand each other's stress points and skill strengths and weaknesses, they need to be trained in language skills and general cultural knowledge. More importantly to develop cultural agility, they need to have substantial interpersonal interaction including social events and participating as a team in analogues and other team training events.

Operational Assessment

From June 14, 2010 to June 16, 2010 the research team visited the Johnson Space Center in Houston, Texas to interview subject matter experts (SME) who were familiar with team training in isolated and confined environments for current missions and the training challenges associate with long-duration missions (e.g., Mars). A total of 16 interviews were conducted. Five of the 16 interviews were conducted via teleconference because of scheduling conflicts. Each interview lasted approximately one hour. The SMEs who were interviewed included astronauts (ASCANs, short- and long duration), flight controllers and managers, training and psychological support specialists in EVA and robotics, SFRM trainers, and psychological support personnel. Interviewees were asked to briefly describe their position and role at NASA and to discuss their involvement with training. Next, the interviewees were asked to discuss current team training at NASA and to consider training needs, emphasis, and methods for long-duration missions. Each interview was recorded and transcribed. Based on a review of the recorded and transcribed interviews, the research team identified content categories. Interview responses were grouped according to these categories. Below, we provide a summary of the main points or "take aways" organized by content category. The interview comments are organized by topic area and are available by request from the authors.

Team Training

NASA has applied team training research results to its current missions by utilizing training to ensure that flight crews, flight controllers, and their interactions result in safe and successful missions. There are a number of training methods used to develop team-related skills, cross-training, and shared mental models in flight crews and flight controllers, and to facilitate effective interaction between the flight crew and various supporting teams on the ground. These include analogues, simulations, table-top simulations, formal courses, virtual reality, and T-38 training. SFRM skills are increasingly recognized as being important for crew safety and mission success. SFRM skills are embedded in training for new ASCAN classes. Crews attend required cross-cultural courses, intensively study Russian language, interact with crew members from other space agencies during ASCAN training, and when a flight crew is assigned they spend time training and visiting with crew members both in the U.S. and in other countries (primarily Russia). Also, as part of formal training and mission debriefs, senior, experienced astronauts are providing coaching, mentoring, and sharing knowledge to help new or less experienced crews obtain both tacit and explicit knowledge gained during previous missions. In addition to team training, psychological support is available to the crew and their familiespreflight, during flight, and during reintegration.

NASA has modified the training flow to ensure that new ASCAN classes develop SFRM skills. SFRM skills are emphasized in courses and integrated into analogues, simulations, and technical training. SFRM training has been approached in a less systematic manner for flight controllers and astronauts from previous ASCAN classes.

Analogues

NASA Extreme Environment Mission Operations (NEEMO). Aquarius is an undersea laboratory used during the NASA Extreme Environment Mission Operations (NEEMO). The base, located several miles off the coast of Key Largo, Florida, is owned by the National Oceanic and Atmosphere Administration (NOAA) and managed by the University of North Carolina. The NEEMO experience places astronauts in an environment with challenges that parallel the hostile physical and stressful psychological environment experienced during long-duration missions. These challenges can include allowing the crew to experience the effects of gravity in space, on the Moon, and Mars, providing a compressed timeline for completing tasks, practicing procedures such as EVAs and emergency procedures used to rescue crew members, and performing tasks with delayed and limited communications with the mission control crew. For example, on May 10, 2010, NASA sent two astronauts, a veteran undersea engineer and an experienced scientist, to Aquarius to learn more about working in an environment that is analogous to space (NEEMO 14). The crew lived aboard the underwater laboratory, ventured out on simulated spacewalks, operated the crane and maneuvered the vehicles much like explorers would in setting up a habitat on another planet. As the crew interacted with these developing technologies, they provided information and feedback to NASA engineers. The crew simulated removing a mockup of the Lunar Electric Rover from the lander, retrieving small payloads from the lander and the ocean floor, and simulated the transfer of an incapacitated astronaut from the ocean floor to the deck of the craft. The rover and lander mockups were similar in size to vehicles NASA is considering for future long-duration missions.

National Outdoor Leadership School (NOLS). NOLS is a remote wilderness expedition involving teaching technical outdoor skills, leadership, and SFRM skills in a stressful, rugged outdoor setting with extreme environmental conditions (e.g., backpacking for two weeks with other astronauts in a cold hostile environment to understand and develop team skills, leadership skills, team dynamics, survival skills, and create an awareness of stressors in self and others).

The informal feedback from astronauts, ASCANs, and trainers about NOLS and NEEMO experiences is uniformly positive. Both NOLS and NEEMO are necessary for crew training and contribute uniquely to mission success. NOLS is a good way for the crew to get to know each other quickly (how they respond to stress, trigger points) so the team can understand its strengths and weaknesses and who can best play specific roles in different types of situations. NOLS is good for helping build teamwork, eroding hierarchies and professional issues by rotating leadership roles (e.g., civilians may artificially place more value on military crew opinions and expertise), and identifying pressure points at the beginning of ASCAN training and/or assigned crew training. NEEMO is a better analogue for long-duration space flight than NOLS because it includes many of the conditions that a crew will face such as living in a small area, experiencing a work schedule that is not under personal control, limited food and supplies, indirect or limited communications with mission control and experts from a distance, and physical hazards. NEEMO is a good introduction to risk elements for individuals who haven't previously flown in space.

Interviewees report that while NEEMO is a good analogue for isolated confined environments, current stays should better simulate long-duration mission conditions. There is a need to evaluate if the current duration of NEEMO is sufficient to elicit conditions and psychological reactions present in long-duration mission (e.g., stress, trigger points, disruption of sleep cycle, and interaction in confined space). That is, there is a need to ensure that NEEMO has the highest fidelity possible to a long-duration mission environment. The interviewees emphasized that whatever analogue is used for long-duration missions, it needs to involve confined space, several awake-sleep cycles, delayed communications, and test the ability of crew resourcefulness. Interviewees expressed concern that the time the crew spends in NEEMO not be lengthened to the point that the experience adds to physically and emotionally exhausting the crew beyond what they will experience as a result of their demanding training schedule for a long-duration mission.

Analogues are useful for evaluating and developing flight crews SFRM skills. The effectiveness of analogues depends on the extent to which the environment parallels the hostile physical and stressful psychological environment that crews will experience on a long-duration mission in a confined space. NOLS is better suited for ASCANs training. It introduces ASCANs to a stressful environment, teaches survival skills, emphasizes the importance of SFRM skills, and helps identify self and other's "pressure points." Assigned flight crews should attend NEEMO because it best represents the psychological and physical conditions found on a long-duration space mission.

ASCANs Training Flow

The training flow for the current astronaut class is new. There is little Shuttle training and the focus has shifted to ISS. The length of the training is 18 to 20 months. After 18-20 months, ASCANs are given a technical position (management) until assigned to a flight. If they are assigned to ISS, they will have to complete an additional 2 years of training before their mission.

ASCAN training starts with NOLS, which emphasizes team building and survival training in a cold climate. Next, the ASCAN class is divided into aviators and non-aviators to fly in a T-38 jet trainer. The ASCANs receive 6 weeks of basic aviation training with the Navy in Pensacola, Florida. Pilots fly in the front seat of the T-38, non-aviators fly in the back seat with a more experienced pilot (e.g., current astronaut) in the front seat. The aviation training includes CRM training and 12 flights. ASCANs also attend training in three flows/areas: ISS system, robotics, and EVA. They receive Russian language training for approximately 5 to 6 hours per week. After graduation in May 2011, ASCANs are ready to be assigned to a mission/flight. Until they are assigned, they are given technical duties such as supporting flight controller training. They also continue T-38 training to maintain certification (100 hrs for pilots, for "backseaters" training hours are needed to become certified as a "backup"), Russian language training, and advanced EVA and robotics skills classes. Starting with the 2009 class, ASCANs have to be proficient in ISS, EVA, and Robotics. This is different from previous ASCAN classes. In previous ASCAN classes, astronauts were assigned to specialize in the specific task (EVA or robotics) they were proficient in. This reflects the differences in Shuttle and ISS missions. Currently in ISS, each astronaut must demonstrate proficiency across ISS, robotics, and EVA. When ASCANs are assigned to a flight, they are assessed for their skills and tasks (formal qualification) to identify what they have retained and what refresher or additional training they need.

The current ASCAN training flow places appropriate emphasis on developing and evaluating team skills, including SFRM skills. However, long-duration missions in confined space introduce new environmental and psychological challenges for effectively using SFRM and other team skills. The ASCAN training flow should include a higher fidelity training experience related to long-duration missions in confined environments (e.g., NEEMO).

Flight Controllers Training Flow

Flight controllers become operators after a year and a half of training and simulation. Astronauts in the unassigned pool assist with these simulations. Flight controllers need to complete three generic simulations per week for certification. In each simulation, there is a checklist of skills and attitudes that must be displayed. The simulations are high fidelity using software simulating all systems and space environment. Data similar to what you would see and in the form you would receive it on a mission is driven to both the crew and flight controllers.

After obtaining operator certification, flight controllers sit console for night shifts and basic operations for the station. They continue to train on more sophisticated operations and gain more experience within a specific area, at which point they become specialists and handle more delicate operations. It is a 2- to 3-year process for each "seat" to get certified. The flight controllers can get certified for multiple seats. The assumption is that,after the flight controllers become certified for multiple seats, they will be able to train other controllers and astronauts. This is a relatively new and motivating aspect of flight controller training.

The rationale for adopting this training flow (instead of the previous “front room/back room” dichotomy) was that efforts to reduce staffing at night and on weekends led to more cross-trained and more experienced flight controllers being assigned to console on weekends, which appeared to an inefficient use of talent. In the new training flow, individuals gain experience on console during quiet times and should be more prepared when they step into more difficult roles (e.g., during EVA or docking). Also, the training flow was changed to make it more appealing by providing quicker certification and new training flows focused on development of specific expertise. The new training flow was developed to reduce the turnover of controllers who were trained by NASA but left for other positions at private firms with government contracts.

The new flight controller training flow helps ensure that NASA retains a sizable pool of talented individuals with multiple certifications. This will give NASA flexibility in determining how to schedule and assign flight controllers for long-duration missions (e.g., number of flight controllers assigned to the mission and length of their work shifts). SFRM experiences for flight controllers appear to be less systematic and institutionalized than for ASCANs. Flight controllers need to receive more opportunities to develop SFRM skills. Also, a needs analysis should be conducted to identify what new skills sets flight controllers need for a long-duration mission.

Space Flight Resource Management Training Program (SFRM)

SFRM is a continually evolving process in crew and flight controller training. SFRM has been the topic of specific courses, and is embedded in other types of astronaut training such as table-top and high-fidelity simulations. Beginning in 2009, every ASCAN class is now required to have an SFRM element. It is a formal "class" and embedded in other training. Formal courses focus on self-care, self-management, leadership, and cross-cultural issues. NOLS, NEEMO, the T-38, the moon-base table-top simulations, and other high-fidelity simulations require use of SFRM skills to deal with normal, preventive, and problem situations. Moon-base simulations work well to elicit SFRM skills related to communications, conflict resolution, and decision making. The use of movies and video clips from previous missions and anecdotes provided by more experienced astronauts in the SFRM class has been useful for making the training meaningful and interesting. Expansion of SFRM into training has had a positive impact on crews, given the constraints that ISS crews are announced at different times, have different training flows, and crew members can be replaced during training. SFRM likely will be more critical in long-duration missions because of the need to deal with the unknown, ground communications delays, an international crew, more confined personal space than on station, and the need to work together to solve new and unfamiliar issues and problems that will occur in a new vehicle heading toward Mars.

Some interviewees suggested that SFRM is an underserved area for flight controllers (as well as astronauts). Little team training is embedded in technical training and SFRM is not mandated for flight controllers . Departments have the discretion to have flight controllers participate in SFRM training but they have limited resources to conduct and support it. However, they recognize that SFRM training is important for flight controllers who operate together as a team and are part of a larger multi-team system (crew, flight controllers) who need to work together to ensure crew safety and mission success. SFRM training with the flight controllers is primarily technical and not done with the crew. This may be due to time constraints. SFRM training involving both the crew and flight controllers is important for developing trust, increasing awareness of each other's roles, helping them learn how to best communicate given the anticipated communication delays on a long-duration mission and understanding each other's verbal and non-verbal behaviors (e.g., what does it mean when a crew member speaks loudly or talks in a low tone of voice?). This is also important because a long-duration mission will represent a significant change in operations from flight control to flight support. Flight controllers and mission control will be unable to be part of "real-time" operations. This is currently not the case with ISS, which is controlled by ground commands. For a long-duration mission, this will be difficult or impossible because of communications delays.

SFRM skills are evaluated by psychologists observing training. For example, debriefs in moon-base simulation from psychologist evaluators focus on SFRM skills. However, the feedback/evaluation piece of SFRM is underdeveloped. There are concerns about the validity of the rubrics and scales used for measurement, the criteria used, and how evaluations are shared and used. There also are concerns about whether the feedback provided is of sufficient quality and quantity to be useful for SFRM skill development.

There is a need to determine whether the current skills emphasized in SFRM are appropriate for long-duration missions in confined space and if additional skill sets should be developed. To help ensure proper assessment and development of SFRM skills, an evaluation of the metrics, scales, and quality and quantity of feedback provided should be conducted.

Long-Duration Flight Issues

A long-duration mission to Mars or an asteroid will pose significant new challenges that have not been faced during current Shuttle and ISS missions. Flight crews will find themselves in a confined space for many months with significant "quiet" time especially during travel to and from Mars. Communication will experience significant delays of twenty minutes or more between the crew and the ground. Crews will have to learn new skills, refresh previously learned skills, and exercise significant autonomy in applying these skills to problems and issues they will face. Thus, we anticipate that long-duration flight will have a significant influence on training issues for both astronauts and ground control.

Cross-Cultural Issues. Similar to ISS missions, a long-duration mission will likely involve an international crew including astronauts from the United States and other nations including Canada, Japan, Europe, Russia, and China. The effectiveness of multicultural teams can be affected by a unique set of issues including cultural differences in communications, decision making norms, adaptation, reactions to stress, conflict management strategies, gender roles, and attitudes towards hierarchy and authority. Currently, assigned crews for ISS may spend limited formal training time together as a crew and the critical training they do receive focuses on specific tasks with limited time to assess and develop SFRM skills. For example, ISS crews meet each other but do not train together until later (in some instances, not at all) in the training flow (i.e., at 18 months crews are involved in full crew event simulations in both the U.S. and

Russia). The amount of time spent training together is based on the crew's travel schedule rather than on requirements to participate, needs assessment, or an evaluation of the crew's SFRM skills. The barriers to more team training include the fact that other countries have their own team training, crews are at different points in the training flow (making it difficult to bring them together), and the costs associated with travel. However, it is important to emphasize that significant efforts are made by the space agencies to get the crews and their families to visit each other in their home countries and socially interact to become familiar with each other's personalities and habits and to help facilitate an understanding of cultural norms and values that may have been emphasized in language training (e.g., Russian language training).

Most emphasis has been placed on U.S. astronauts understanding Russian language and culture because the Soyuz is the primary vehicle used by astronauts to reach the ISS and the U.S. and Russia are the two most advanced space programs in the world with a history of joint missions. There also are differences in U.S. and Russian training approaches and philosophies, which are larger than the differences with the other international space agencies whose astronauts participate in the ASCAN training flow (e.g., Japanese Space Agency (JAXA), European Space Agency (ESA), Canada Space Agency (CSA)). These differences include:

1. Russians fit their behavioral training into the technical training, parachute, and survival training.
2. Russians emphasize observation and evaluation in last two simulations and while in space, pay is based on performance of tasks (e.g., last two simulations are evaluated by a "commission"). U.S. astronauts also receive feedback from instructors, trainers,

psychologist observers, but it is used to identify weaknesses and areas that need further training.

3. Russia does not send their cosmonauts to ASCAN training unlike astronauts from ESA and CSA.
4. Russian training tends to be more theoretical. Crew members are responsible for taking notes and more directly responsible for learning with less documentation provided (e.g., books, technical manuals).
5. Extensive psychological support is provided to U.S. astronauts, but not Russian cosmonauts. Russian cosmonauts are less likely to ask for support, perhaps because it will negatively affect their evaluation and pay.

Language competency has affected both Russian cosmonauts and U.S. astronauts. Russian cosmonauts are not assigned EVA and robotics tasks on ISS because of language difficulties. For the Soyuz vehicle, U.S. astronauts who struggle with language cannot succeed in “left seat.” The “left seat” requires flight engineering skills needed to control critical systems and thus requires a complex understanding of the Russian language to understand procedures, panels, displays, switches, etc. Great strides in trust and communications with Russia have been made over the years through joint missions (Skylab, Shuttle, ISS). The cosmonauts and astronauts do get to know each other on a personal level and develop SFRM skills through travel and simulation training in U.S. and Russia. Also, more senior astronauts from both the U.S. and Russia have provided useful insight into the personal and cultural nuances that crew members can expect to experience while on a joint mission.

Team training and participation in analogues for assigned crews is especially important because of the need to understand the “trigger points” of other crew members, which may be difficult to understand because of cultural differences. Understanding of these “trigger points” is important for the crew to be able to exchange roles to capitalize on their strengths and weaknesses to deal effectively with certain situations, develop a common language, understand non-verbal communications (e.g., gestures), and recognize how crew members deal with conflict and stress. It is important to emphasize that issues that arise between crew members may result from both individual style differences and/or cultural differences (e.g., autocratic leadership style). Team training experiences in high-fidelity analogues and simulations can help crew members understand both individual and cultural differences and develop cultural agility.

Changes in Skills and Mindset. All of the interviewees suggested that there are several new types of skill sets and changes in mindset that crew members and flight controllers need for a long-duration mission. First, for the crew, they will be faced with the need to cope with loneliness and boredom that they will be especially vulnerable to during the anticipated 6-month trip to and from Mars. This continues the evolution of crew time being completely scheduled on space missions to the crew having more autonomy to schedule and complete most tasks. For example, on Shuttle the crew is always busy with ascent, decent, and experiments. ISS crews are occupied while en route to and from the Space Station aboard the Soyuz but, once on board, the work pace is slower, free time is available, and they have time to exercise and work by themselves. A long-duration mission to Mars will give the crew much more free time on a vehicle than previous crews have experienced on Shuttle and ISS, most likely in a habitat providing less personal space. There will be a need for both private space for crew members as well as a common gathering space. Crew members will need to find some designated personal space to get away and psychologically recharge. It is also important that the crew not be given “busy work” to cope with boredom but instead be provided with experiments and tasks that are mission focused. The crew should be involved in value-added, mission-related work (i.e., meaningful work) while enroute to and from Earth on a long-duration mission. This could include training, mission planning, tactical planning, work on assignments, completing debriefs, and analyzing data. The concept of a “job jar,” currently used on ISS missions also could be useful.

A second issue that the crew and flight controllers will face is a communication delay of twenty minutes or more. This has several implications. First, the crew will have to be self-reliant, autonomous, self-sufficient, and unable to rely on the ground. The crew will have to troubleshoot problems and take action, consulting experts as the second step only if the situation allows. Second, the crew will have to be trained in general principles and have on-board expert systems and simulations. Currently, crews for ISS missions can train for specific task and skills proficiency right up to launch. Refresher training for ISS is limited to emergency drills and reentry training. The exact specifications for EVAs and landing are well known. Proficiency due to skill decay is not an issue. For long-duration missions, training will have to focus on the dynamic parts of flight with refresher training on board.

Crews and their families currently receive excellent psychological support beforelaunch, during the mission, and in reintegration when they return from their mission. From a self-care perspective on a long-duration mission, crew members will need to establish some type of communication with Earth and their families and friends, even if it is asynchronous communications such as prerecorded audio or video. A major issue will be how to deal with the effects of long-term confinement. Astronauts will need to understand the psychological effects of long-term confinement and be encouraged to ask for support from the ground when they need it. Also, materials related to personal hobbies and activities and projects that keep the crew "healthy" will have to be carefully identified and included on the vehicle before launch because they will not be able to be sent on a resupply spacecraft.

Recommendations for Practice

Based on the operational assessment and the literature review, we make the following recommendations for team training for long-duration missions:

1. Assigned long-duration mission crews should spend time together in high-fidelity analogues such as NEEMO to ensure that the crewmembers are aware of "stressors" or "pressure points" and the crew develops a "common language", trust, SFRM skills (e.g., resourcefulness), and guided self-direction needed to successfully execute the mission. The crew should spend sufficient time in the analogue to experience stressors such as several asleep-awake cycles, each other's personal habits, and reduced personal space. The exercises that take place in the analogue should mimic, to the extent possible, emergency situations and day-to-day operations that the crew will encounter on a long-duration mission including communication delays, having to troubleshoot and fix problems that require use of general principles and on-board training systems, and switching roles to maximize team success and minimize individual personal and skill weaknesses. NEEMO is especially valuable for

assigned flight crews because the situations they will encounter will require them to shed routinized, rigid interaction patterns and prepare them to adapt to emerging crisis situations that may occur on a long-duration mission to Mars.

2. The NOLS and NEEMO analogues are popular, but are perceived to be elective and not formally and explicitly required in preparation for space flight. Currently, NOLS is recommended for Shuttle, NOLS and NEEMO are optional for Station. The current ASCAN classes attend NOLS but assigned crews do not have to attend an analogue. Given the challenges that crew members will face on long-duration missions (e.g., stress, need to be resourceful, need to identify team members trigger points to facilitate effective role exchanges), it is especially important that the use of both NOLS and NEEMO be continued. Because of its high fidelity with long-duration mission conditions, there needs to be an institutional requirement for assigned crews to attend NEEMO. This institutional requirement also is important because many current astronauts have not received the same level and type of SFRM training as newer ASCAN classes. This would allow experienced astronauts to further strengthen their SFRM skills and share their explicit and tacit knowledge about their missions with the less experienced crew members. Also, requiring assigned crews to attend NEEMO is a recommendation supported by research, which has found that the stability of team membership moderated the relationship between team training and team outcomes (Salas et al., 2008). Intact teams that underwent training improved the most on process and performance outcomes.

3. Complete a review with emphasis on time spent in different types of training: in this instance, the length of the training flow is not the most important issue, the most

important issue is “Is what I am learning necessary?” Unnecessary training depletes crew emotional resources and strains family and other relationships because they have to continuously travel to and from home.

4. Evaluate the amount of time spent on certain types of training involving extreme or unusual conditions: e.g., emergency situations on take off and landing that would be catastrophic for the crew, robotics involving grappling, drawing blood. Astronauts reported that they felt that it was not useful to go through every exact extreme scenario that could occur. Rather, they preferred a focus on general principles that they could apply to extreme or unusual circumstances. This is especially important for long-duration mission because critical tasks can be completely or partly trained for before the mission with refresher training or remaining training modules provided on board the vehicle. Training should involve more general principles (mechanics, troubleshooting) because the crew will have to take primary responsibility for fixing urgent problems (e.g., equipment failures that could compromise crew safety). Because of communication delays, a long-duration mission crew will not have the luxury of contacting the ground and waiting for experts to respond. Contacting experts on the ground is possible for less important problems, but the crew will still have to deal with long delays in communications.

5. Team training needs to be based on the concept of operations. On long duration missions, the crew will be out of touch with ground control. The crew will have to know tasks and time constraints and be given more tactical control (What tasks have to be completed at a certain time versus at the discretion of the crew?). Clear guidelines need to be provided to the crew and training made available using a JIT

tool. The culture of mission control would have to evolve from ISS; Mission Control needs to become mission facilitation. This change in philosophy especially differs from Shuttle missions, which have a full working schedule with crew activity controlled by the ground.

6. Long-duration crews will have to be self reliant. Resourcefulness is a current training objective that will increase in importance for long-duration missions. This means the crew needs to have support tools and know how to apply them in the correct situations. Simulations, analogues, and emergency response training can be used to teach resourcefulness by having the crew use what is available and apply general principles to solve problems. Also, a long-duration crew must be capable of dealing with psychoses because even normal, well-adjusted crew members can experience them after long periods of isolation. This is especially important because ground support will be limited in the psychological support that they can provide during a Mars mission due to the time lag in communications.
7. Training time should be devoted to dynamic situations that could compromise safety and mission. These dynamic situations need to be identified. Key questions need to be addressed such as: What tasks and skills need to be trained early with repetitive practice? Which tasks and skills can receive less repetition but competence established with refresher training before flight and/or on board the vehicle using data packages, and high-fidelity virtual reality (immerse yourself in task before performing it) or simulators?
8. Flight controllers will have to be trained on how to deal with the time delay, keeping mission safety the first priority and mission success the second. Flight controller

staffing and scheduling will need to be reviewed. For example, one issue that needs to be considered is: Should the current schedule with three shifts of seven or eight teams rotated be maintained or should fewer teams with the "best" flight controllers be used through the entire mission?

9. Both research on MTS and the operational assessment suggest the need to enlarge the "team" involved in training to include controllers, crew, and all personnel involved in mission. On successful ISS missions, crew and flight controllers have interacted together to build trust and understand each other's reactions to communications. It is especially important that the multi-team system (crew and flight controllers) develop and maintain high levels of SFRM skills given the conditions on a long-duration mission.

10. Because of the importance of SFRM skills for long-duration missions, more frequent evaluations of crew member skills need to be provided from the perspectives of peers, instructors, and trainers.. Mentoring and coaching resulting from these evaluations should be framed positively, i.e., not seen as improving a weakness but rather helping to sharpen a strength. One model that should be considered is that of executive coaching. Psychologists should be involved before a team or individual crisis occurs. The focus should be on diagnosing and preventing SFRM problems and issues rather than "fixing" then after they occur.

11. Coaching, mentoring and various types of interpersonal relationships and knowledge sharing strategies are currently being used. The interviewees' view of these is uniformly positive because of both the tacit and explicit knowledge that new and less experienced astronauts can learn from more "expert" colleagues. The use of peer-to-

peer, expert-to-peer and other types of relationships should be formalized within the astronaut program. This would also increase meaningfulness of training and engagement of crew. Also, it would be useful if “lessons learned” from previous ISS missions can be formally documented and shared with astronauts, flight controllers, trainers, and instructors. This information should be summarized in a way that protects the identity and preserves the confidentiality of the crew members. This information also needs to be formalized to augment current astronaut debriefs because the types of tacit and explicit knowledge of ISS crews may be invaluable to helping crew members preparing for long-duration missions anticipate issues, obstacles, and develop the resourcefulness they need to cope with the uncertainty they will encounter.. This information could be accessible through a knowledge management systems similar to The Center for Army Lessons Learned that the U.S. Army has established for sharing lessons learned in the battlefield (see http://usacac.army.mil/cac2/call/index.asp).

12. The results of the ongoing “Mars 500” study should be carefully followed and reviewed. This study can help us better understand team training needs. Questions being investigated in this study include adaptation, group structure, and communications of confined and isolated crews, determining the effects of group dynamics and loneliness on cognitive and emotional adaptation to extreme, confined environments and the implications of personal values for interpersonal compatibility and individual adaptation during a long-duration mission. NASA should conduct similar type of studies of team effectiveness and team processes using crews in the NEEMO analogue.

13. Crewmembers on long-duration missions will need cultural agility. For long duration missions, it is imperative that crewmembers' levels of cultural agility can be assessed so they are aware of it, and they are assigned to flight crews early enough so they can together share training and team-building experiences such as NEEMO, NOLS, and simulationsto better assess and develop their task work and teamwork skills, understand crewmembers' stress points, develop backup behaviors, and develop cultural agility that will be critical for the success of long-duration missions. The current emphasis that NASA places on language training and country-specific knowledge is important, but not the only prerequisite for cross-cultural agility. The development of cross-cultural agility also depends on significant peer-to-peer contact with persons from different cultures and opportunities to question personal assumptions and realize the cultural limits of personal knowledge and behavior.

Research Recommendations

Teams and Team Effectiveness

For long-duration missions, there will be a greater need to ensure that team training allows for greater development of task work and especially teamwork. This includes equipment, work procedures and strategy, awareness of member responsibilities and role interdependencies, understanding of team member's preferences and skills.

1. Most studies of team mental models have examined either teamwork or task work content. The relationship between teamwork and task work content has not been established. Research is needed to examine the interactions between types of mental models. For example, the benefits of the teamwork mental models may depend on whether the task work mental model is shared. Is the same or longer time needed to

develop teamwork and task work mental models? What are the implications on performance of a crew that has not fully developed a teamwork model but has a fully developed task work model? What are the implications of experiences like NEEMO and NOLES for the development of shared teamwork and task work models and development of cultural agility? Research is needed to examine other indicators of team effectiveness such as cohesion and psychological safety. What are the implications of shared teamwork and task work models for crew effectiveness in multiteam systems like NASA (flight controllers, astronauts)?

2. A fundamental assumption of the team mental model literature is that greater sharing of knowledge among team members results in increased team effectiveness. Is this the case? What roles require greater convergence of knowledge? In which roles are complementary or distributed roles better? What team mental model content domains should converge? Would a team mental model consisting of distributed task work knowledge and overlapping teamwork knowledge result in higher performance? In what cases do team mental models result in a dysfunctional crew? F or example, too much similarity across member models may result in inaccurate views that are validated by other team members rather than ignored or discarded. Do certain team norms (such as constructive confrontation norms) moderate the relationship between teamwork and task work model similarity and performance?

3. The interviews we conducted suggest that while many of the tasks and the knowledge, skills, abilities, and other requirements of a long-duration mission will be similar to those encountered in Shuttle or on the ISS, there will be significant differences. A team task analysis is needed to identify the job and task requirements that a team will encounter on

a long-duration mission. A thorough team task analysis involves a requirements analysis, identifying the specific tasks that compose the target job, identifying the teamwork taxonomy (team work behaviors frequently performed or needed in team performance situations), coordination analysis, select relevant tasks for training, translation of KSAs or competencies that will become the target for selection, training, and development, and link the KSAs back to the team tasks (see Burke, 2004). The result of the team task analysis can identify which types of competencies or KSAs are required for teamwork on a long-duration mission. According to Cannon-Bowers et al. (1995), these include context-driven competencies specific to both the task and the team, which are best developed through practice with actual team members in realistic task environments (e.g., vehicle simulators, NEEMO), task contingent competencies specific to the task but not the team and can be trained with or without actual teammates (e.g., EVA, robotics), team-contingent competencies, which are specific to the team but not the task requiring a training environment including actual team members across a variety of tasks (e.g., SFRM) , and transportable competencies, generic to both the task and the team, which can be trained using a variety of tasks and team members (e.g., T-38 training).

4. The operational assessment suggests that on a long-duration mission in a confined physical space, knowledge of team members' pressure points and stressors is critical and back-up behaviors may be especially important. Backup behaviors include providing a team member with feedback or coaching, assisting a team member in carrying out a task, and taking charge of and completing a task for a teammate. It is especially important for team members to understand when a teammate is overloaded or experiencing some other factor (e.g., stress, boredom, loneliness) that decreases their performance and deploy their

own resources to help the struggling team member. Despite conventional wisdom that backup behavior is a critical team characteristic, few empirical studies have evaluated the role of backup behavior in team effectiveness, and Barnes et al. (2008) suggest that backup behaviors might actually encourage negative social behaviors such as social loafing, dependence, and neglect of personal task work. Research is needed to identify the types and taxonomy of backup behaviors that are related to effective crew performance so that team training in analogue environments and simulators can create problems and issues that would evoke team backup behaviors.

5. Research is needed to establish the relationship between cross-level understanding, shared mental models, and team performance. Huber and Lewis (2010) suggest that a high level of cross-level understanding is related to high team learning and performance regardless of the degree of convergence of the team's mental model. A high level of cross-understanding allows members to discuss their differences and perspectives and also mitigates the negative impact of discussion bias favoring commonly held information (a bias that can occur when mental models are shared). A high level of cross-understanding might help explain why diversity of opinions and ideas would be expected to have a negative effect on group cohesion and performance, but it does not.

6. The crew members on a long-duration mission are also part of a larger multi-team system that includes flight controllers and other experts on the ground. The larger multi-team system can be considered a virtual team because the members must coordinate their work through asynchronous electronic communications. It is estimated that communications between the crew and the flight controllers during a long-duration mission will be delayed for at least twenty minutes. Driskell et al. (2003) proposed four

processes to be especially important for virtual teams: cohesiveness, status, counternormative behavior, and communications. Bowers et al. (2008) notes that trust, collective efficacy, and team orientation may all be affected by distance. The current research on virtual teams focuses on processes and attitudes that may affect interaction between team members. Research is needed that examines the processes and attitudes that are important in virtual multiteam systems, which will be the reality of long-duration missions.

7. Research is needed to understand how cultural agility influences crew effectiveness. For example, how does each crew member's cultural orientation (minimalist, adapter, integrator) influence team processes such as coordination, communication, back up behavior and overall crew effectiveness. How is cross-cultural agility developed within a flight crew and how do crews adapt and integrate cultural differences to effectively deal with normal and abnormal situations? Do analogue experiences such as NOLS and NEEMO improve crew member's cultural agility?

Team Training and Training Systems

On long duration missions it is not possible to train for every possible scenario or problem ahead of time or to maintain skills that were trained on the ground but may require refresher training due to long time periods of lack of use. The training requirements will continue to evolve from Shuttle to ISS to long duration missions. Training for certain aspects of the mission such as ascent and decent will likely continue be highly scripted but due to the novel equipment and new and unexpected circumstances, problems, and challenges that will be encountered during long duration missions, the crew training on the ground will need to have a greater emphasis on general skills and principles that they can apply to a wide variety of tasks. Also, due to communication delays for urgent problems there may not be sufficient time for astronauts and mission control on the ground to learn about, practice and/or model how to deal with the situation and then communicate it to the crew (as was the case for the astronauts on the Expedition 24 mission who were faced with an EVA to replace a failed ammonia pump module). On-board individual and team training and decision-support systems are needed for long duration missions to provide refresher training for the skills acquired on the ground, to learn new skills, and to understand how to apply general skills to specific tasks that occur during the mission. These training and decision-support systems could be embedded in flight systems and be specifically designed for supporting on-board training and decision support.

1. Research is needed to identify the dynamic situations that could compromise safety and mission success on a long duration mission. Care must be taken to identify the tasks and skills that need to be trained early with repetitive practice, and those which can receive less repetition but provide a refresher right before flight and/or onboard the vehicle using data packages, virtual reality (e.g., immerse yourself in task before performing it). Also, which part, if any, of training should occur on the ground and which should be left for on-board training, also known as expanding and progressive training?

2. Research is needed to identify skill decay patterns in order to determine which skills are best trained on the ground and/or in-flight and the level of refresher training that needs to be provided. For example, which learned skills have a slow or gradual decay pattern compared to skills which have a quick and steep decay pattern. How do instructional characteristics (e.g., variable practice, active versus passive learning,

contextual interference) influence skill decay patterns? What is the best training schedule for skill maintenance? How could refresher training needs be determined? Based on the task, the skills, or both?

3. Although the *potential* advantages of technologies such as simulations, virtual reality, virtual worlds, iPads, and hand-held devices for learning and delivery of instruction have been touted, little research has examined their effectiveness or what features of instruction should be included to maximize learning and transfer. Research is needed to identify if these devices are most effectively used as primary instructional devices, as part of a blended learning approach, or to facilitate transfer of learning by providing skill and knowledge refresher training. Also, the instructional features that need to be included (or deemphasized) in these devices needs to be determined. For example, does the novelty of experiences in a three-dimensional world help trainees recall the experience but interfere with retention and transfer of training? Does the ability to synchronously or asynchronously collaborate with expert peers, instructors, or mentors enhance learning beyond just providing learner control?

4. There is a need to evaluate the effectiveness of SFRM training. Research is needed to collect data linking SFRM training to improved mission performance using naturalistic observations of crew interactions (process) and mission performance (outcomes) during table-top and high fidelity simulations and analogues such as NEEMO. Also, using training records and instructor evaluations, ratings, and written comments it is possible to assess whether the crew meets the required SFRM proficiency level for training events related to the mission profile. The naturalistic observations and instructor evaluations

should continue to be collected and reviewed and used for improving SFRM stand-alone and embedded training.

It is important to note that the research issues we have identified above are unique, but have some overlap with those identified by Barshi (2009). Barshi (2009) identified other important research questions related to training philosophy, methods, and content, training delivery, and vehicle interface and design that should be addressed to enhance the effectiveness of long-duration missions.

References

*Alonso, A., Baker, D. P., Holtzman, A., Day, R., King, H., Toomey, L., Salas, E. (2006). Reducing medical error in the military health system: How can team training help? *Human Resource Management Review*, *16*(3), 396-415.

Ball J.R., Evans E.H. Jr. (Eds.) *Safe passage: astronaut care for exploration missions*. Washington, DC: National Academy Press; 2001:136 –71.

Baker, D. P., Gustafson, S., Beaubien, J. M., Salas, E., & Barach, P. (2003). *Medical teamwork and patient safety: The evidence-based relation.* Washington, DC: American Institutes for Research.

Barlow, D. 1996. Health care policy, psychotherapy research and the future of psychotherapy. *American Psychologist* 51:1050–1058.

Barshi, I. (2009). Training and decision support systems. In M.K. Kaiser & J.McCandless (eds.) *Space Human Factors Engineering Technical Gap Analysis White Papers* (pps. 55-78). NASA Ames Research Center: NASA (Technical Report 2009-214580).

Beck, A.T. 1993. Cognitive therapy: past, present and future. *Journal of Consulting and Clinical Psychology* 61:194–198.

Beck, A.T., and G.Emery. 1985. *Anxiety Disorders and Phobias: A Cognitive Perspective.* New York: Basic Books.

Bowers , C., Smith, P.A., Cannon-Bowers, J.A., & Nicholson, D. (2008). Using virtual worlds to assist virtual teams. In P. Zemliansky (ed.), *The handbook of research on virtual workplaces and the new nature of business practices* (pp. 408-423). Hershey, PA: IGI Publishing Group.

*Brandl, J, & Neyer, K (2009). Applying cognitive adjustment theory to cross-cultural training for global virtual teams. *Human Resource Management*, 28 (3), 341-353.

Burke, C.S. (2004). Team task analysis. In N.A. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (eds.), *Handbook of human factors and ergonomic methods* (pp. 365-423). Boca Raton, FL: CRC Press.

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

Caligiuri, P. (2010). *Managing the global workforce* (Wiley).

Cannon-Bowers, J.A. & Bowers, C. (2011). Team development and functioning. In S. Zedeck (ed.), *APA Handbook of Industrial and Organizational Psychology* (vol. 1) (pps. 597-650). Washington, DC: American Psychological Association.

*Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: recent findings from applied research. *Current Directions in Psychological Science*, *7*(3), 83-87.

Cannon-Bowers, J., Salas, E. & Converse (1993). Shared mental models in expert team decision-making. *Individual and group decision making- current issues*, 221-246.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Voipe, C. E. (1995). Defining competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco: Jossey-Bass.

Cautela, J.R., & Ishaq, M. 1996. *Contemporary Issues in Behavior Therapy: Improving the Human Condition.* New York: Plenum.

Chen, G.-M., & Starosta, W. J. (1996). Intercultural communication competence: A synthesis. In B. Burleson (Ed.), *Communication yearbook* Vol. 19. (pp. 353–383). Thousand Oaks, CA: Sage.

* Clement, J.L., Boyd, J.E., Kanas, N., & Saylor, S. (2007). Leadership challenges in ISS operations: Lessons learned from junior and senior mission control personnel. *Acta Astronautica*, 61, 2-7.

Clement, J.L., and Ritsher. JB (2005) Operating the ISS: cultural and leadership challenges. In 56th International Astronautical Congress, Fukuoka, Japan: International Astronautical Federation 2005.

*Daniel, L., & Simpson, E. (2009). Integrating team training strategies into obstetrical emergency simulation training. *Journal for Healthcare Quality: Promoting Excellence in Healthcare*, *31*(5), 38-42.

*Delise, L.A., Gorman, C.A., Brooks, A.M., Rentsch, J.R. & Steele-Johnson, D. (2010). The effects of team training on team outcomes: A meta-analysis. *Performance Improvement Quarterly*, 22 (4), 53-80.

*DeChurch, L.A. & Marks, M.A. (2006). Leadership in multiteam systems. *Journal of Applied Psychology*, 91, 311-329.

*DeChurch, L.A. & Mesmer-Magnus, J.R. (2010). The cognitive underpinnings of effective teamwork. *Journal of Applied Psychology*, 95, 32-53.

*Dion K.L. (2004). Interpersonal and group processes in long-term spaceflight crews: Perspectives from social and organizational psychology. *Aviation, Space, and Environmental Medicine*, 75, C36-C43.

Driskell, J., Radtke, P., & Salas, E. (2003). Virtual teams: Effects of technological mediation on team performance. *Group Dynamics: Theory Research and Practice*, 7, 297-323.

*Earley, C. P., & Peterson, R. S. (2004). The elusive cultural chameleon: Cultural intelligence as a new approach to intercultural training for the global manager. *Academy of Management Learning and Education*, *3,* 100-115.

*Ellis, A.P.J., Bell, B.S., Ployhart, R.E., Hollenbeck, J.R., & Ilgen, D.R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology*, 58, 641–672.

*Flin, R., & Martin, L. (2001). Behavioral markers for crew resource management: A review of current practice. *International Journal of Aviation Psychology*, 11(1), 95-118. doi: Article

* Fox, Johnson, Gagliano, Passarello, Moore, Resurreccion and Reed (2006). Team training: A tool to improve resident teamwork and communication. *Journal of the American College of Surgeons, 209*(3), S113.

*Gelfand, M.J., Erez, M. & Aycan, Z. (2007). Cross-Cultural Organizational Behavior. *Annual Review of Psychology*, 58, 479-514.

Goldstein, I.L. & Ford, J.K. (2002). *Training in organizations* (4th ed.). Belmont, CA: Wadsworth.

Gourinat, Y., Apel, U., & Delbart, F. (2010). The AEROCREW mission: Training space session at Ny Ålesund Arctic Base. *Acta Astronautica*, *66*(1/2), 74-77.

Gudykunst, W. B. (1998). Applying anxiety/uncertainty management (AUM) theory to intercultural adjustment training. *International Journal of Intercultural Relations*, 22(2), 227–250.

Hamman, W. R. (2004). The complexity of team training: what we have learned from aviation and its applications to medicine. *Quality & Safety in Health Care [Qual. Saf. Health Care]*. *Vol. 13, 13*, i72-i79.

*Helmreich, R. L. (2000). Culture and error in space: implications from analog environments. *Aviation, Space, and Environmental Medicine*, *71*(9 Suppl), A133-9.

*Heinrichs, W. L., Youngblood, P., Harter, P. M., & Dev, P. (2008). Simulation for team training and assessment: Case studies of online training with virtual worlds. *World Journal of Surgery, 32*(2), 161-170.

*Hoegl, M., Weinkauf, K., & Gemuender, G. (2004). Interteam coordination, project commitment, and teamwork in multiteam R & D projects: A longitudinal study. *Organization Science*, 15, 38-55.

Holland, A.W., L.Looper, and L.Marcondes-North. 1993. Multicultural Factors in the Space Environment: results of an international shuttle crew debrief. *Aviation, Space and Environmental Medicine* 64:196–200.

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

*Huber, G.P. & Lewis, K. (2010). Cross-understanding: Implications for group cognition and performance. *Academy of Management Review*, 35, 6-26.

Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and trust in global virtual teams. Organization Science, 10(6), 791–815.

*Kanas N. (2004). Group interactions during space missions. *Aviation, Space, and Environmental Medicine*, 75, C3-C5.

*Kanas, N., Sandal, G., Boyd, J. E., Gushin, V. I., Manzey, D., North, R., Leon, G. R., Suedfeld, P., Bishop, S., Fiedler, E.R., Inoue, N., Johannes, B., Kealey, D.J., Kraft, N., Matsuzaki,I., Musson, D., Palinkas, L.A., Salnitsky, V.P., Sipes, W., Stuster, J., Wang, J. (2009). Psychology and culture during long-duration space missions. *Acta Astronautica*, 64(7-8), 659-677

*Katz-Navon, T., Naveh, E., & Stren, Z. (2009). Active learning: When is more better? The case of resident physicians' medical errors. *Journal of Applied Psychology*, 94: 1200-1209.

Kealey D.J. (2004). Research on intercultural effectiveness and its relevance to multicultural crews in space. *Aviation, Space, and Environmental Medicine*, 75, C58-C64.

Kelley, A.D., and N.Kanas. 1994. Leisure time activities in space: a survey of astronauts and cosmonauts. *Acta Astronautica* 6:451–457.

Klein, C., Salas, E., Burke, C.S., Goodwin, G.F., Halpin, S.M., Diazgranados, D. et. al (2006). Does team training enhance team processes, performance, and team member affective outcomes? A meta-analysis. *Academy of Management Proceedings*, 1-6.

Kozlowski and Ilgen (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the public interest,* 7(3), 77-124.

Kozlowski & Klein (2000). From Micro to Meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, 3(3), 211-237.

*Kraft, N. O., Lyons, T. J., & Binder, H. (2003). Intercultural crew issues in long-duration spaceflight. *Aviation, Space, And Environmental Medicine, 74*(5), 575-578.

*Kraus, D., & Gramopadhye, A. K. (1999). Team training: Role of computers in the aircraft maintenance environment. *Computers & Industrial Engineering, 36*(3), 635.

Langan-Fox, J. (2003). Skill acquisition and the development of the team mental model: An integrative approach to analyzing organizational teams. In M. West, D. Tjosvold, & K.G. Smith (eds.), *International handbook of organizational team work and co-operative working*: 321-360. London: Wiley.

Lazarus, A.A. 2000. Multimodal replenishment. *Professional Psychology: Research and Practice* 31(1):93–94.

* LePine, J.A., Piccolo, R.F., Jackson, C.L., Mathieu, J.E., & Saul, J.S. (2008). A metaanalysis of team work processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61, 273-307.

*Lerner, S., Magrane, D., & Friedman, E. (2009). Teaching teamwork in medical education. *The Mount Sinai Journal of Medicine, New York, 76*(4), 318-329.

*Lim, B.-C, Klein, K. (2006) Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, 27(4) 403-418.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331-352

*Marks, M.A. DeChurch, L.A., Mathieu, J.E., Panzer, F.J., & Alonso, A. (2005). Teamwork in multiteam settings. *Journal of Applied Psychology*, 90, 964-971.

* Marks, M.A., Mathieu, J.E. & Zaccaro, S.J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*,26. 356-376 .

*Marks, M.A, Sabella, M.J., Burke, C.S., & Zaccaro, S.J. (2002). The Impact of Cross-Training on Team Effectiveness. *Journal of Applied Psychology*, 87 (1), 3A–13.

Mathieu, J. E., Marks, M.A., & Zaccaro, S.J. (2001). Multi-team systems. In N. Anderson, D. Ones, H.K. Sinangil, & C. Viswesvaran (eds.) *International handbook of work and organizational psychology* (pps. 289-313). London: Sage.

McComb, S.A. (2007). Mental model convergence: The shift from being an individual to being a team member. In F. Dansereau & F. Yammarino (eds.), *Research in multilevel issues* (vol. 6) : 95-147. Amsterdam: Elsevier.

Mohammed, S. & Dumville, B.C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior*, 22, 1-58.

Mohammed, S, Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, 36 (4): 876-910.

Noe, R.A. (2010). *Employee training and development* (5th ed.). New York: McGraw-Hill/Irwin.

Nullmeyer, R.T. & Spiker, V.A. (2003). The importance of crew resource management behaviors in mission performance: Implications for training evaluation. *Military Psychology*, 15 (1), 77-96.

*Orasanu, J. (2005). Crew collaboration in space: A naturalistic decision-making perspective. *Aviation, Space, and Environmental Medicine*, 76, B154-B163.

Osland, J. S. (2000). The journey inward: Expatriate hero tales and paradoxes. *Human Resource Management*, 39(2&3), 227–238.

*Palinkas, L. A., & Suedfeld, P. (2008). Psychological effects of polar expeditions. *Lancet*, *371*(9607), 153-163.

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

Power, K.G., R.J.Simpson, V.Swanson, L.A.Wallace, A.T.C.Feistner, and D.Sharp. 1990. A controlled comparison of cognitive-behavior therapy, diazepam and placebo, alone and in combination, for treatment of generalized anxiety disorder. *Journal of Anxiety Disorders* 4:267–292.

*Ritsher, J. (2005). Cultural factors and the international space station. *Aviation, Space, and Environmental Medicine*, *76*(6), B135-B144.

Rosen, J.B., and J.Schulkin. 1998. From normal fear to pathological anxiety. *Psychological Review* 105:325–350.

Salas, E., Burke, C. S., & Cannon-Bowers, J. A. (2002). What we know about designing and delivering team training: Tips and guidelines. In K. Kraiger (Ed.), *Creating,* *implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 234-259). San Francisco: Jossey-Bass.

Salas, Bowers, and Cannon-Bowers (1995). Military team research- 10 years of progress. *Military Psychology*, 7(2), 55-75.

*Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S.M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors*, *50*(6), 903-933.

Salas, E. Dickson, T.L., Converse, S.A., & Tannenbaum, S.I. (1992). Toward an understanding of team performance and training. In R.W. Sweezy & E. Salas (eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.

Salas, E., Nichols, D. R., & Driskell, J. E. (2007). Testing Three Team training strategies in intact teams: A meta-analysis. *Small Group Research*, *38*(4), 471-488.

Salas, E., Stagl, K.C., Burke, G.C. & Goodwin, G. (2007). Fostering team effectiveness in organizations: Toward an integrative framework. *Nebraska Symposium on Motivation*, 52, 185-243.

*Sandal, G. M., & Manzey, D. (2009). Cross-cultural issues in space operations: A survey study among ground personnel of the European Space Agency. *Acta Astronautica*, *65*(11-12), 1520-1529.

*Shapiro, M. J. (2004). Simulation based teamwork training for emergency department staff: does it improve clinical team performance when added to an existing didactic teamwork curriculum? *Quality and Safety in Health Care*, 13(6), 417-421. doi: 10.1136/qshc.2003.005447

Smith, E. Ford, J.K., & Kozlowski, S.K. (1997). Building adaptive training expertise: Implications for training design strategies. In M.A. Quinones & A. Dudda (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pps 89-118). Washington, DC: American Psychological Association.

Smith-Jentsch, K.A., Cannon-Bowers, J.A., Tannenbaum, S.I. & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research*, 39, 303-327.

*Stachowski, A.A., Kaplan, S.A. & Waller, M.J. (2009). The benefits of flexible team interaction during crisis. *Journal of Applied Psychology*, 94, 1536-1543.

TeamSTEPPS (2005). Agency for Health Care Research and Quality. Retrieved February 2010 from http://teamstepps.ahrq.gov/taq_index.htm

Tomi L., D. Kealey, M. Lange, P. Stefanowska, V. Doyle. Cultural challenges facing the International Space Station personnel: results of a cross-cultural training requirements definition survey. In: A. Vakoch (Ed.), *Psychology of Space Exploration: Contemporary Research in Historical Perspective*, (Chapter 9), in press.

Wolpe, J. 1958. *Psychotherapy by Reciprocal Inhibition.* Stanford, Calif.: Stanford University Press.

Yu, T., Sengul, M., & Lester, R. H. (2008). Misery Loves Company: The Spread of Negative Impacts Resulting From an Organizational Crisis. *Academy of Management Review*, *33*(2), 452-472.

* Notes articles included in the literature review

Appendix 1: Journals included in the review.

Academic Emergency Medicine: Official Journal of The Society For Academic Emergency Medicine
Academy of Management Learning and Education
Academy of Management Proceedings
Academy of Management Review
Acta Astronautica
Administrative Science Quarterly
Annual Review of Psychology
Applied Ergonomics
Aviation, Space, and Environmental Medicine
CA Magazine
Communications of the ACM
Computers & Industrial Engineering
Contemporary OB/GYN
Current Directions in Psychological Science
Ergonomics
Human Factors
Human Resource Management Review
Human Resource Management
International Journal of Aviation Psychology
International Journal of Industrial Ergonomics
Journal for Healthcare Quality: Promoting Excellence in Healthcare
Journal of Applied Psychology
Journal of Applied Social Psychology,
Journal of European Industrial Training
Journal of Organizational Behavior
Journal of the American College of Surgeons
Lancet
McGill Journal of Medicine
National Productivity Review (Wiley)
Organization Science
Organization Studies
Performance Improvement Quarterly
Personnel Psychology
Quality & Safety in Health Care
Safety Science
Small Group Research
The Mount Sinai Journal of Medicine
Theoretical Issues in Ergonomics Science
Training & Development Journal
World Journal of Surgery