OFFICE OF THE DIRECTOR OF ADMINISTRATION AND MANAGEMENT
1950 DEFENSE PENTAGON
WASHINGTON, DC 20301-1950

Ref: 10-A-1065-A1
FOID: 10-F-1065

Mr. Steven Aftergood
Federation of American Scientists
1112 16th Street NW #400
Washington, DC 20036

Dear Mr. Aftergood:

This responds to your July 29, 2010, Freedom of Information Act (FOIA) appeal. You appealed the decision of the Freedom of Information Division (FOID) in response to your May 26, 2010, FOIA request.

After carefully considering your appeal, and as a result of discussions between the Office of the Director, Defense Research and Engineering (DDR&E), FOID and this office, I am releasing 92 pages of this document to you in part, a copy of which is enclosed. Portions of this document remain exempt because they are protected from disclosure under the FOIA pursuant to 5 U.S.C. § 552(b)(6), which concerns material the release of which would constitute a clearly unwarranted invasion of the personal privacy of third parties.

If you are dissatisfied with my action on your appeal, the FOIA permits you to file a lawsuit in federal district court in accordance with 5 U.S.C. § 552(a)(4)(B).

Sincerely,

CHUNG.JOO.Y.1512306507

Digitally signed by CHUNG.JOO.Y.1512306507
Date: 2021.02.25 16:17:09 -05'00'

Joo Y. Chung
Director
Oversight and Compliance

Enclosure:
As stated

cc:
FOID

~~FOR OFFICIAL USE ONLY~~

Quest for Truth: Deception and Intent Detection

Study Leaders:
S. Keller-McNulty

Contributors:

R. Breaker
N. Lewis
S. Cazares (IDA Intern)
P. Dimotakis
F. Dyson
D. Eardley
R. Garwin
M. Goldberger
R. Jeanloz
G. Joyce
W. Press
M. Ruderman
R. Schwitters
J. Sullivan
J. Vesecky
P. Weinberger
E. Williams

October 2008

JSR-08-143

FOR OFFICIAL USE ONLY

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**		
1. REPORT DATE *(DD-MM-YYYY)* October 2008	**2. REPORT TYPE** Technical	**3. DATES COVERED** *(From - To)*
4. TITLE AND SUBTITLE Quest for Truth: Deception and Intent Detection		**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) S. Keller-McNulty		**5d. PROJECT NUMBER** 13089022 **5e. TASK NUMBER** PS **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The MITRE Corporation JASON Program Office 7515 Colshire Drive McLean, Virginia 22102		**8. PERFORMING ORGANIZATION REPORT NUMBER** JSR-08-143
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Human Performance Modeling & Simulation Division Night Vision Lab 10221 Brubeck Road Ft. Belvoir, VA 22060		**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution authorized to US Government and Contractors; Administrative/Operational Use; October 29, 2008. Other requests for this document shall be referred to NRO.		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT This report is in response to Department of Defense (DOD) interest in the detection of deception and intent to do harm. The report provides an assessment of the potential utility and efficacy of monitoring and assessing human behavioral neurophysiology and verbal and nonverbal communication to determine human intent in a military context. The JASON study focused on the process that would need to be implemented in order to identify covert combatants and ultimately infer intended actions is outlined. A key finding in this report is the need to establish a discipline of science in the development and deployment of potential technologies, including interrogation methods, that have been proposed to be useful in this setting. This report gives recommendations regarding how DOD can remedy this deficiency.		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (b)(6)
a. REPORT FOUO	b. ABSTRACT FOUO	c. THIS PAGE FOUO	UL		19b. TELEPHONE NUMBER *(include area* (b)(6)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

~~FOR OFFICIAL USE ONLY~~

Contents

1 INTRODUCTION 5

2 PROBLEM STATEMENT 9

3 CULTURE 13

3.1 Cultural Separation of Societies: What can we learn from the Greeks and neighboring cultures? 14

3.2 Navigating Cultural Variances 16

3.3 Cultural Impact 17

4 TRAINING 19

4.1 Field Training and Red Teams 19

4.2 Technology Insertion 22

4.3 Virtual Environments for Training 23

5 TECHNOLOGY 25

5.1 Heart Rate, Heart Rate Variability, and Respiration Rate ... 26

5.2 Remote Technologies for HR, HRV, and RR 29

5.3 Thermal Imaging for Fear and Deception Detection 36

5.4 Other Sensors 40

5.5 Future Attribute Screening Technologies Program 43

6 INTERVIEW METHODOLOGIES 49

6.1 Informal Interviewing 50

6.2 Cognitive Interview 53

6.3 Technology Assisted Interviewing 55

6.4 Cultural Impact on Ability to Detect Deception and Infer Intent 56

7 ESTABLISHING A SCIENCE BASE: LEARNING THE TRUTH 59

7.1 The Science Base 60

7.2 The Path to Improvement 62

7.3 Some Proposed Full-Scale Trials 65

7.4 Goals and Metrics Matter: A note on rare events 65

8 CONCLUSIONS 67

APPENDIX A **69**

FOR OFFICIAL USE ONLY

Abstract

This report is in response to Department of Defense (DOD) interest in the detection of deception and intent to do harm. The report provides an assessment of the potential utility and efficacy of monitoring and assessing human behavioral neurophysiology and verbal and nonverbal communication to determine human intent in a military context. The JASON study focused on the process that would need to be implemented in order to identify covert combatants and ultimately infer intended actions is outlined. A key finding in this report is the need to establish a discipline of science in the development and deployment of potential technologies, including interrogationmethods, that have been proposed to be useful in this setting. This report gives recommendations regarding how DOD can remedy this deficiency.

EXECUTIVE SUMMARY

The JASON was asked by the Department of Defense (DOD) to provide an assessment of the potential utility and efficacy of monitoring and assessing human behavioral neurophysiology and verbal and nonverbal communication to determine human intent in a military context. In addition to evaluating current capabilities, The JASON was asked to delineate potentially useful research topics that could further progress in these areas for critical national security applications.

Problem Space: The process for identifying a covert combatant and then assessing intent is described by the following steps.

1. It is necessary to be able to gain an understanding of the baseline environment, i.e., what is *normal.* This could be normal crowd behavior at the airport, typical street interactions in a village, or normal physiological response of an inmate.
2. It is necessary to be able to recognize an anomaly in the context of thebaseline environment. It may be possible to induce a shock into the baseline that could help anomalies present themselves.
3. The anomaly must be engaged. This could be through an interviewor interrogation, possibly assisted with technology. The technological assistance may be remote or in contact with the individual, and covert or overt. The polygraph is an example of an overt contact technology for interrogation.
4. An assessment of current and future behavior needs to be made. In amilitary context it is important for the assessment to delineate desire, intent, and decision to act.

~~FOR OFFICIAL USE ONLY~~

Findings

- Scientific evidence exists to support identification of individuals using DNA, iris scans, and possibly fingerprints.
- Incident-specificpolygraph tests can discriminatelying from truth telling at rates well above chance, but such methods are not useful for screening.
- Quoted success for identifying deception and intent are *post-hoc* and incorrectly equate success in identifying covert combatants with the identification of drug smugglers, warrant violators, etc.
- Detection of deception is confounded by culture.
- No compelling evidence yet exists to support remote monitoring of physiological signals in an operational scenario.
- No scientific evidence exists to support the detection or inference of future behavior, including intent.
- Training exists in DOD for developing intuition and understanding of the adversaries culture and environment.

Overarching Recommendation

- Train the soldier to be aware of diverse cultural environments and complex human dynamics.
- Look for small incremental advances that add value to the well-trained soldier, integrating the soldier throughout the research and development processes to avoid unrealistic expectations.
- Establish a discipline of*science* for technology and methodology development and deployment.
- Develop valid performance and effectiveness goals and metrics.

- Develop clear and measurable expectations.
- Link the research, development, and deployment to the problem space, including clear definition of the operational scenarios.
- Use *unbiased* peer-review for program plans.

- Leverage research expertise in biomedical sensor development

1 INTRODUCTION

The JASON was asked by the Department of Defense (DOD) to provide an assessment of the potential utility and efficacy of monitoring and assessing human behavioral physiology and verbal and nonverbal communication to determine and forecast human intent. In addition to evaluating current capabilities, the JASON was asked to delineate potentially useful research topics that could further progress in these areas for critical security applications.

The human dimension has become an increasingly important component of warfare in the 21st Century. Today, the soldier must be equipped to understand new cultures and complex behavioral dynamics. This is clearly articulated in a recent U.S. Army Training and Doctrine Command report
[1].

> *"While unable to rule out conflict with western peer nations, the most likely future clashes will be against opponents that will approach warfare from radically different perspectives that do not conform to U.S. or Western practices. They will view American moral, political, and cultural values as vulnerabilities to exploit. Future conflict will remain savage and bloody with potential horrific attacks on the U.S. and its allies. The Army will face an unconstrained enemy empowered rather than limited by technology. Such adversaries' objectives would not be to destroy or defeat U.S. or allied formations by force of arms, but to shatter political and popular will to continue protracted conflict.*
>
> *Future conflict therefore is likely to usher in another era of small "savage wars of peace." Rather than retreat, the U.S. and its military forces, often with allies and other interested nations, will remain engaged in complex power struggles worldwide in order to protect national interests. For Army forces, a strategy of engagement places a greater premium on understanding the human dimension.*

> *While there have been recent and profound changes in the ways and means to conduct war, its essence will not change. Conflict will remain complex and chaotic, and human frailties and irrationality will continue to characterize war's nature. Just as in the past, the root causes of future conflicts will arise from fear, hatred, greed, honor, and ambition. The human dimension of war defies simple logic. For that reason, understanding as much as possible about this dimension becomes critical to influencing and achieving favorable outcomes in future conflicts.*
>
> *War will continue to be primarily a contest of opposing wills. Ambiguity, danger, physical exertion, friction, and chance, constitute the "climate of war" which contribute to the "fog of war" with which commanders must contend in future operations. Technology, intelligence, and operational design can reduce uncertainty; however, commanders still must make decisions based on incomplete, inaccurate, or contradictory information. These factors will continue to play a predominant role in the environment of future full spectrum operations."*

An ultimate goal is to train and technologically support the soldier for understanding, detecting, and forecasting human intent. This presents an enormous challenge for the DOD due to the complexity of human behavior, as well as the cultural variances in the ways a person can express his or her intent and the inducements leading to implementation of that intent.

DOD feels that while a development effort that strives to make progress in this area is a high-risk undertaking, the potential impact of any substantive progress can be of great value for a number of important applications. This is especially relevant to combat situations where identifying combatants and/or terrorists is a high priority. Therefore, insight into how context, culture, body language, verbal expression, and human physiologic changes could provide measurable signatures of human intent could add significant value to the national security community.

FOR OFFICIAL USE ONLY

DOD tasked the JASON to provide a scientific evaluation and assessment of the following:

- *"What are the most promising methods for developing relevant behaviors/signatures of terrorists/covert combatants?*
- *What are the most promising techniques for measuring, detecting, characterizing and identifying behaviors/signatures relevant to covert combatant behaviors/signatures?*
- *To what degree can human behavioral neurophysiology and verbal and nonverbal communications taken together improve our ability to assess human intent?*
- *Can virtual environments serve as an effective methodological and training tool for elucidating the factors that indicate human intent? • Can virtual environments be used for training advanced sensor skills to make soldiers more sensitive to the enemy's tactical behavior and natural communications occurring in a civilian environment?"*

The report begins in Section 2 with a formal statement of the problem goal, which is to detect the intent to do harm in a military stetting. This will be contrasted with the problem of detecting deception. Section 3 puts culture into context for the human dimension of inferring intent. Section 4 discusses methods for training the soldier to identify hostile intent. Section 5 discusses sensor technologies for detecting hostile intent. Section 6 discusses interview methodologies for confirming the intent to do harm. Section 7 draws attention to the need for fostering the use of scientific best practices in human dimension research. Section 8 will draw some final conclusions.

2 PROBLEM STATEMENT

An ultimate goal for the DOD is to detect the intent to do harm of a covert combatant. This means one needs to be able to identify and interact with the combatant for the purpose of assessing *intent*. The combatant may be in a crowd at an airport or train station, in a remote village, or being detained in a cell. The combatant may be from a foreign culture, i.e., a culture that is foreign to the observer. The combatant, from his or her perspective, may be skilled in deception and be on a holy or culturally justified mission versus a harmful mission. The problem is hard because different people with intent to do harm will show different states of stress, nervousness, and anomalous behaviors.

The briefers, listed in Table 1, were consistent regarding the steps in the process for identifying a combatant and then assessing intent. These steps are given here.

1. It is necessary to be able to gain an understanding of the baselineenvironment, i.e., what is *normal*. This could be normal crowd behavior at the airport, typical street interactions in a village, or the normal physiological response of an inmate.

2. It is necessary to be able to recognize an anomaly in the context of thebaseline environment. It may be possible to induce a shock into the baseline that could help anomalies present themselves.

3. The anomaly must be engaged. This could be through an interviewor interrogation, possibly assisted with technology. The technological assistance may be remote or in contact with the individual, and it could be covert or overt (the polygraph is an example of an overt contact technology for interrogation).

4. An assessment of current and future behavior needs to be made. In amilitary context it is important for the assessment to delineate desire, intent, and decision to act.

Table 1: Briefers

Barbara OKane, PhD (Night Vision & Electronic Sensors Directorate and Badguyology™ Program Manager)	US Army RDECOM CERDEC
Paul Ekman, PhD	UCSF and Paul Ekman Group
Jay Nunamaker, PhD	University of Arizona and Group Systems
Matthew Jensen, PhD	University of Arizona
Jeffrey McNeil, PhD	United States Army Special Operations Command (USASCO)
Troy Brown, PhD	Defense Academy for Credibility Assessment (DACA)
Dan Martin, PhD	MRAC, LLC
Robert Burns	HSARPA
Sgt. Peter DiDomenica	Massachusetts State Police
Lt. Buzz Benson	Gwinnett County Sheriff's Department, Atlanta, GA
Stacy Marsella, PhD	University of Southern California (USC)
Louis-Philippe Morency, PhD	University of Southern California (USC)
Jeremy Bailenson, PhD	Stanford University
Anton Leuski, PhD	University of Southern California (USC)

Each part of the process has complications. The human *normal* state is highly dependent on culture and context. At a minimum, significant training is needed to gain intimate familiarity of new environments, frequently nonWestern environments. Virtual worlds and serious computer games, coupled with field training, could be helpful components in the development of a cultural understanding of one's relevant surroundings.

Identifying an anomaly or anomalous behavior within the background is also difficult. Given the low expected frequency of covert combatants, many false-positive identifications should be expected. Several new security screening programs were presented by the briefers. Data should exist to help understand the rate of false-positive identifications in airports or train stations where these methods are being employed by thousands of screeners. The JASON did not have access to data or peer-reviewed information regarding the claims of success of the various security screening programs. However, anecdotally it was found through media reports that in recent airport screenings, 1-2% of the individuals selected for secondary screening through the Transportation Security Administration's Screening Passengers by Observation Technique program, a non-profiling methodology, were arrested (e.g., http://seattlepi.nwsource.com/local/344868_airportprofiler26.html). It was stated that these arrests were for a variety of charges, including possession of drugs, weapons violations, and outstanding warrants and not, however, for terrorist identification and detention, the reasons for which the secondary screening was performed (e.g., http://blogs.computerworld.com/node/ 6971) . Metrics for assessing the success of such processes would be required to decide their efficacy and efficiency. For example, would the outcome have been different if 1-2% of individuals walking about in a city in the U.S. were randomly selected for secondary screening?

Detecting, let alone identifying, intent is simply not feasible today through technology, interview techniques, or coupled interview and technology solutions, even within the same culture. Figure 1 summarizes the state-of-the-art in confirming identity, detecting deception, and inferring intent. JASON [2] concluded that identity could be inferred through DNA, iris scans, and possibly fingerprints, provided adequate cross-linked and indexed databases to match this information were available. JASON [2] and National Research Council [3] concluded that the polygraph, with a well trained examiner, could detect incident specific deception. Both studies indicated more research would be needed to enable the development of a clandestine polygraph equivalent and the development of methods for the detection of more general deception. The National Research Council [3] report did not find polygraph techniques useful or hopeful for screening of other forms of deception and intent.

JASON [2] and National Research Council [3] encouraged exploration into remote biometric measurement techniques as potentially useful means to enhance interview technologies. However, the reports were clear that such advances would likely stem from the biomedical community and would need extensive scientific testing to assess if relevant signatures could be obtained. In the present context, obtaining these signatures could assist the soldier in confirming anomalies in the baseline environment and assessing the anomalies' current and future intent.

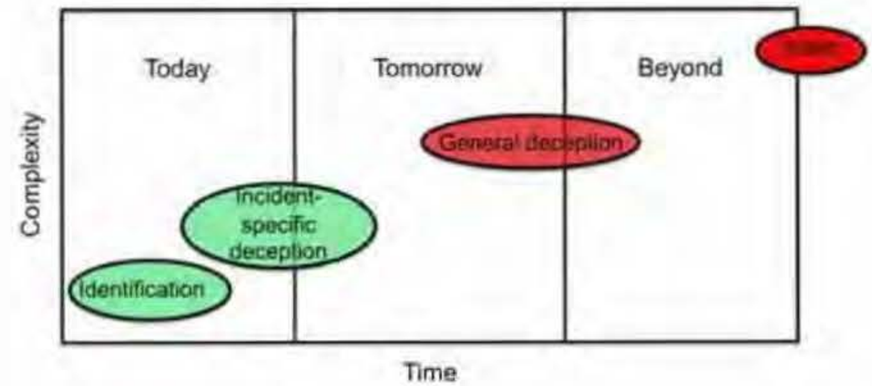

Figure 1:

JASON [2] and National Research Council[3] felt there was some indication in the literature that demeanor detection could be accomplished by learning to recognize facial expressions, verbal and non-verbal cues, emblems, and body movement. However, both studies indicated the absence of an experimental science base to validate these hypotheses. Intelligence Science Board [4] covers some of the most important challenges that exists with the problem of assessing intent of a covert combatant, namely culture barriers.

3 CULTURE

A useful step toward inferring intent is to understand what is normal in the surrounding environment. What is normal has huge cultural nuances. The nuances go well beyond linguistics and dress code. The famous quote from Sun Tzu [5] is appropriate here.

> *"If you know the enemy and know yourself, you need not fear the result of a hundred battles."*

The problem is that today's soldier may not be prepared to know either!

Findings:

- Detection of deception is confounded by culture.

Recommendation:

- The soldier's cross-cultural understanding of his adversary must be developed and must go well beyond situational awareness training in order to enable the detection of anomalies in a baseline environment. • DOD should continue (increase) training in foreign language skills because it helps sensitize the soldier to the cultural differences they will need to confront and accommodate.

It is essential to understand cultural variances in the expression and implementation of intent in a military context. Many people discuss the importance of culture. There has been a volume of research done on the theoretical development of social networks, frequently *post facto*. What has not been done is the translation of this social understanding into useful and validated methods for the soldier on the ground. Some studies have been designed to evaluate cultural differences on detecting anomalies and deception. However, they fall short of reality, typically involving university students and largely focusing on linguistic differences. The reviews by Hazlett and Kleinmann in [4] put much of this into context.

The JASON believes that this important area has not received the serious attention it merits. This may be because what has been written in this area is not sufficiently practical for the reader to understand and use to recognize the deep rooted cultural differences that surround all of us in daily life. This may also be because the explicit and implicit recommendations are not easily reduced to practice. When it comes to the detection and assessment of covert combatants intent, the consequences of cultural differences may be insurmountable without years of mentoring, training, and education.
To that end, the JASON will present the cultural case in a unique fashion.

3.1 Cultural Separation of Societies: What can we learn from the Greeks and neighboring cultures?

Culture leaves important, often indelible, imprints on human behavior. Relevant to this study, these imprints include the patterns of response to stimuli of various kinds, the ability and propensity to deceive as necessary, how people formulate tactics and strategy, the likelihood of detection and detectability of intent and intended actions by various means. Such behavioral patterns (traits) comprise the unwritten, timeless assumptions that enable the formation and efficient function of societies. To illustrate this point, examples and contrasts with the modern "West" will be drawn from ancient and modern Greece, the Eastern Mediterranean, and the Middle East.

Ancient Greeks valued and cherished the ability to live together in cities and societies. Scorn was attached to solitary existence. When a civilian was punished in ancient Athens, even to death, he or she had the simple option of either staying in Athens and taking the punishment, or simply leaving the city, forbidden from returning and forced to live among barbarians. Socrates had the option of leaving the city after he was convicted of corrupting the city's youth, i.e., teaching mores and values outside the norm. Socrates chose to stay and drink the hemlock.

Words that exist in both Greek and English illustrate the general appreciation of Socrates' notion. *Idiot* in ancient Greek means someone who lives by himself. There is no ancient (or modern) Greek word for privacy because, in ancient times, people who lived by themselves were assessed incapable of functioning with others. By extension, the ancient Greeks perceived problems with a person that needed "privacy". Contrast this with the American idealization of the "lonesome cowboy": the ancient Greeks would have considered him to be an *idiot*. However, in American culture, privacy is (presumed) to be a self-evident, constitutionally granted, and a protected right, even though the word does not appear anywhere in the constitution.

The word *civilization* in Greek is *πολlτlσμζ*, which means, "that which (only) occurs in a city." The word for *funny* in ancient Greek is *αστειοζ*, which literally means "(an attribute only a) city person (can possess)". Non city dwellers are not capable of *humor* (the vapor that a *human* exudes), which has no value unless there are others around to appreciate it.

The ancients both celebrated and strengthened the differentiation between societies as a necessary catalyst for a smoothly operating society. This differentiation was also a means to set societies apart from others that may be in competition. This both helped protect societal identity and allowed enemies to be demonized.

To summarize, the primal need to defend and separate societies leads to the development of cultural behavioral traits that are the product of intentional upbringing and education. That leads, in turn, to deep-rooted differences between cultures. For example, variances in language and dialect, while originally a natural outgrowth of geographic separation, were also valued as a mark of ethnic identity. If the goal is deception, it is difficult to succeed as a cross-cultural impostor, in a culture sufficiently different from ones own, language differences aside.

3.2 Navigating Cultural Variances

Many people who have not deeply experienced the great variance between cultures naturally assume that either others are like them and share the same values, or, worse, that others should share the same values, religion, etc. By extension, many people believe that they have the right to impose their values on others. Observing and interpreting behavior is naturally done using ones own cultural metrics.[1]

As an example, there is a surprising commonality that persists today between the culture found in Crete, Sicily (where Greek was spoken in some

[1] In this case, neglecting the ancient Greek advice, "Do not judge others' (actions/affairs) by yours (your own metrics).

villages until relatively recently), and Corsica[2]. Corsicans today consider themselves as descendants of the ancient Greeks and most Corsican geographic names are traceable to Greek and Latin/Italian, not to French. Corsicans do not consider themselves French and ridicule the French. Conversely, the French do not appreciate the difficulty and insist that Corsicans should behave as the French do. As long as this short-sightedness persists on both sides, the (serious) problems between Corsica and France will continue.

Among these three closely interrelated island cultures, the ability to deceive is developed from a young age, prized, and considered an art. In Western Judeo–Christian cultures, deception and lies are considered a sin, a fact claimed to have been chiseled in stone in the Ten Commandments. Deception is, however, a common behavior among many Eastern-Mediterranean and Middle-Eastern cultures. To many Judeo-Christians, Arabs, for example, are liars. The significant misunderstanding, however, is that what is prized in those cultures is not lying, per se, but the *ability* to deceive. In many Eastern-Mediterranean and Middle-Eastern societies, one is not allowed to lie to people in ones own social network. The Commandment regarding lying actually attests to this: "Thou shalt not bear false witness *against thy neighbour.*" In the original Greek translation of the Old Testament, the phrase for *neighbor* is *the person close to you*, with an ambiguity that is clearly intentional.

Lying within one's own social network is a very serious transgression in Eastern-Mediterranean and Middle-Eastern societies and stigmatizes one's family. Such societies are guided by an honor code, expect their members to abide by many powerful written and unwritten rules, and may need few policemen. One finds few policemen in Crete or Corsica, for example. Contrast this with modern Western cultural and behavioral practices, in which societal systems are designed in the expectation that people will do what they can get away with. The difference in incarceration rate between the U.S. and Europe, especially in Southern Europe, is sobering.

[2] The island of "beauty" is a literal translation of the ancient Greek name for the island

The notion of a social network in these societies is a complex concept which depends on what constitutes the network's unifying principle, *in each case*. For family matters, *depending on what these may be*, the social network may be the immediate or the greater family itself. If to respond to a dispute between factions, the social network is automatically the faction. If the issue is topical, the social network may be the village society. If there is a national threat, the social network is the nation, even though the latter is a relatively modern, externally imposed, concept in the Middle East, leading to many misunderstandings in the West. Briefly, there is no such thing as *the* network, a point that has been missed by efforts to map *the* network and attach individuals to *the* network. This should be borne in mind in the attempts to map such networks with a goal of deriving inferences and predictions from them. The social network is necessary for social *function*, in such societies. While a sinister subnetwork may exist or be formed, as necessary, its formation is part of the same semi-automatic process that adapts and defines *a* network to respond to *a* need, as discussed above.

3.3 Cultural Impact

Understanding cultural variances and nuances, even in one's own culture, develops with maturity. This can pose difficulties for our young soldiers.

The DOD must be realistic in its expectations regarding the soldiers' capacity to become quickly immersed and accepted in different cultures. Substantial training and additional education, e.g., languages, will be needed. New technologies developed to assist the soldier in the identification of covert combatants and the assessment of intent to do harm must be vetted and validated in cross-cultural scenarios.

To further underscore the importance of a healthy cultural cognition, we close with a personal story from a JASON who lived in Tehran for many years. In Tehran, matters are made more complex by the fact that in these cultures, a person maintains a public and a private persona. The former is shared openly, while the latter is reserved for the closed environment. This separation of the close-environment face from the public face is also reflected in the architecture of these cultures from ancient times. The family house had one small door and possibly no windows to the outside, with openings of the living quarters to a surrounded courtyard that served as the family nucleus center, where the hearth was. The ancient Greek word for *hearth* (*εστια*) is the same as the word for *focus*, as in optics. Similarly, in Latin. It was and is considered a failure to allow the private persona to be revealed in an inappropriate environment, and something that approaches a crime, in some cultures, for a woman to do so. Such a separation is considered duplicitous and dishonest in Western cultures.

4 TRAINING

Anecdotally, human endeavor is typically improved by education and training, whether it is some athletic skill, public speaking or rattling off the multiplication tables. Training is needed to understand and prepare for the human dimension in today's and tomorrow's conflicts. In the military context reviewed here, training is needed to both understand and gain experience with diverse cultural environments, to detect deviations from normal behavior in those environments, and to initiate interactions targeted at assessing intent to do harm.

Finding:

- Training exists in DOD today to assist in developing intuition and understanding of adversaries' culture and environment.

Recommendations:

- The development of training programs to prepare the soldier for diverse cultural environments and complex human dynamics may be DODs best near-term investment.

- Red teams, embedded in military training exercises, could be used to field test new technologies and interrogation strategies.
- Integrating the "end users" into technology development and deployment may help identify promising technologies at an early stage and help minimize unrealistic expectations regarding the utility of the technology.

4.1 Field Training and Red Teams

The Defense Science Board (2003)[6] stressed the current importance of the development of red teams and training involving red teams, given the fact that current adversaries are more difficult targets for intelligence than during the Cold War. As was described in Section 3, many of the cultural dynamics the soldier may need to face is counter intuitive to the social dynamics with which he or she is familiar. This training, at a minimum, prepares the soldier for the type of environment they will encounter. It could also prepare them to gain an intuitive understanding of what is normal in that environment and how to detect deviations from normal.

General Meigs [7] characterizes the training involving red teams as the "contact scrimmage" before the game. In reference to Fort Irwin training[3], the Operations Group does everything it can to replicate the actual environment in which the unit will find itself. This effort includes the use of an "enemy", a unit that, in peacetime, does nothing but play Red Force for the training units. This exclusive focus ensures that the Opposing Force is very hard to beat and that they can adapt to the training needs of various units. The Operations Group also takes from the unit commander (or his boss) the collective training skills the unit must master in order to accomplish the mission in the

[3] The Army operates a very realistic training program at Fort Irwin, in the Mojave Desert. In this program, the Army tries to teach their soldiers and officers how to handle themselves in Iraq. Army units come to Fort Irwin just before deployment and spend one to two weeks in training. There is a similar Marine training center at Twenty Nine Palms.

area of operations. They weave these tasks into the exercise. Anomalies are created within events in the exercise to force the unit's chain of command to adapt to the unexpected. Units must confront all of the capabilities that a thinking enemy would likely throw at them.

Normally, a rotation in opposed-force maneuver at a Combat Training Center lasts approximately 10 days. About every other day, the Operations Group conducts After Action Reviews. These are seminars in which the leaders of the units relive the events of the exercise to date, explaining in their own words what went well and what did not. The psychological imprinting of the lessons is very powerful. General Meigs [7] figuratively calls this "scar-tissue learning." In addition to the After Action Reviews that occur at intervals during the exercise, one takes place at the end of the maneuver that wraps things up and summarizes to the unit leadership those tasks and skills that they must continue to improve. This kind of competitive training is one of the unique strengths of the US Army and is believed to be one of the reasons units have done as well as they have in Iraq and Afghanistan.

The Fort Irwin experience will be described in some detail here. At Fort Irwin, surrogate Iraqi villages have been constructed with materials and architecture intended to mimic what is actually encountered in Iraq. For example, buildings have flat roofs, designed for sleeping in hot weather, but from which "snipers" can target troops. The streets are strewn with trash that is typical of that found in Iraq. The debris is used to conceal improvised explosive devices (IEDs), booby traps, etc.

The villages are populated by Iraqi-American actors, dressed as Iraqis with flowing robes for the dignitaries, women with or without veils, teenagers in sneakers and blue jeans. All speak Arabic with a strong Iraqi accent. The troops who have come back from Iraq say that the actors are very realistic, both in appearance and mannerisms.

For most of the troops, the training involves very little technology and is focused on how to interact with people from another culture. This ranges from troop behavior in a hostile crowd of villagers to how their commanding officer interacts with the village mayor. It is important to note that the trainers here are the red team.

Fort Irwin has an IED "laboratory" where versions of IEDs found in Iraq are manufactured for use in training exercises and are emplaced to harass the "trainees." The laboratory personnel, mostly technically inclined young enlisted men, are approximately the same age as their real counterparts in Iraq who are making IEDs to kill the troops. Efforts are made to mimic the actual operating procedures now understood in Iraq. In addition to the bomb makers, there are financiers, less skilled people who emplace the IED's, triggermen, cameramen, etc.

To date, the best report of the value of the Fort Irwin training has been that most officers and soldiers report that they were grateful for having gone through the training soon after they arrived in Iraq. Some of them have gone through the training cycle several times.

4.2 Technology Insertion

The question arises regarding the use of technology in these training exercises. The *Mark 1 eyeball* is considered a useful technique for detection fo IEDs. In the case of the training just described, some may confuse the *Mark 1 eyeball*, i.e., the human observer, with physical technology. In fact it is not. The *Mark 1 eyeball* is a behavior that the soldier can learn. Soldiers are taught to refine their ability to look and scan an area carefully. Subtle differences among the soldiers have emerged in the *Mark 1 eyeball* training. Some soldiers are good at finding IEDs, others not so good.

Now consider true physical technologies. Good new technology is recognized as such by the users in Iraq, and field commanders soon demand as much as they can get. Other technology sits in crates in forward operating bases after battle-hardened troops convince themselves that it gives them negligible protection, and perhaps even subjects them to more casualties because of its overhead and distractions.

Perhaps not all discarded technologies are really useless, and perhaps they suffer because they do not come with a good training team to show the troops how to use the new equipment. However, it seems that if the "endusers" were more integrated into the technology development and deployment process, promising technologies would be quickly identified and poor technologies diverted. In other technology areas, best practices for technology development considers the end-user as the focus of the development (Kortum 2008) [8].

4.3 Virtual Environments for Training

Part of the task statement asked the JASON for an assessment of the use of virtual environments for training to help elucidate factors indicating human intent and for advanced sensor technology use. The JASON did not focus on these components of the task statement and does not have specific findings and recommendations in response to these components of the task statement. This is because the technical challenges, both hardware and algorithms, in virtual environments out-strip any potential current use.

Virtual environments may one day be useful to supplement training, in ways similar to how simulations and serious games have become useful.[9] Facilities are being developed (e.g., http://vhil.stanford.edu/) which allow people to view and independently control computer-modeled images, called avatars, in computer-generated environments. When immersed in such a "Virtual Reality," (VR), the scene and activities experienced within it can change in response to the avatars.[4] Such VR facilities and technologies may have promise as instruments for training of military personnel. One

[4] One outside group which might be fruitfully consulted about this consists of classical psychoanalysts together with some of their long psychoanalyzed patients. For nearly a century such analysts have been involved with personality projections and "transference" phenomena. The psychoanalysts have typically spent many hundreds of hours with each of many patients. It might be interesting to see how well psychoanalysts would predict responses of each of their patients to a special set of VR experiences. If the psychoanalysts predict better then the VR designer group, what are the additional insights psychoanalysts use? How useful is the very long period of intimate contact with the patient in making their prediction more reliable?

advantage would be that training in VR of many scenes similar to those likely to be experienced later in the real world, including interactions with deceivers and covert combatants, could be reproduced and widely distributed.

One can envision a range of VR uses, some that would be accommodated by simple 2D VR systems and others that would require a 3D environment to achieve the desired goals. High-fidelity VR systems that create the most complete 3D experiences will demand large amounts of processing power and communication bandwidth, which are expected to improve over time. These are needed to interpret various user inputs (e.g. body motions or even speech) and provide real-time imagery, sound, and even tactile environments back to the user.

One of the key challenges to developing VR technology for training soldiers to detect hostile intent will be the ability to simulate many virtual people and to capture and evolve group behaviors in a virtual world. A near term goal might be to take a very specific problem like the *Mark 1 eyeball* and develop and validate a virtual scenario that better trains, or complements the training, of the soldier in a Ft. Irwin type environment. A long term goal would be to provide a virtual training environment which could provide results comparable to the training at Fort Irwin. This goal is judged to be extremely difficult, and currently available VR capability falls short of achieving it in many ways. A medium-term goal would be a VR humvee driving simulator, with area-specific roads crowded cities, and IED threats, for training of drivers before deployment to a specific operational area. Vehicle simulators have long been successful for training of aircraft pilots and tank drivers (as at Fort Knox); the goal here is training of humvee drivers to recognize cues for IEDs or hostile vehicles.

Appendix A reviews some of what the JASON learned about virtual environments. DOD may find it useful to follow some of these technological advances. As with all DOD technology adoption, rigorous scientific validation and demonstration should be applied. In particular, while visually or otherwise appealing, metrics would need to be defined against which VR technology utility would be assessed.

5 TECHNOLOGY

Inspired by the success of the Recognitions of Combatants–Vehicle (ROCV) program [10], DOD created the Badguyology™ program to explore a similar paradigm in deception and intent detection. The ROC-V concept uses visual cues of geometric shape and hot spots from thermal imagery to discriminate between blue versus red combat vehicles. Training the soldier to recognize the shape cues and use the thermal technology has virtually eliminated fratricide of armored vehicles.

Moving from the identification of vehiclesto human deception and intent will not be easy or quick. Promising technology will need to be vetted and validated, scientifically. It will need to be demonstrated that the technology adds measurable value to the soldier in the defined operational scenarios. Technology fielded too soon will neither save lives nor benefit the soldier.

Findings:

- No compelling evidence yet exists to support remote monitoring of physiological signals in an operational scenario.
- No scientific evidence exists to support the detection or inference of future behavior, including intent.

Recommendations:

- Establish a discipline of *science* for technology and methodology development and deployment.
- Leverage research expertise in biomedical sensor development.

This section briefly covers several technologies that have been proposed for the measurment of physiological responses under stress. Upon reviewing these technologies, the JASON was disappointed to find that even the first steps of validating these technologies has not yet been documented in the published, peer-reviewed literature.

5.1 Heart Rate, Heart Rate Variability, and Respiration Rate

Physiological signals change in response to both physical and emotional stress due to the influence of the autonomic nervous system[11]. Sensors have been developed to measure these changes. Traditional sensors require contact with the subjects skin and some are the basis of the standard polygraph [3]. In contrast, novel sensors have been proposed for the *remote* monitoring of physiological signals [12, 13, 18, 19]. Due to their non-contact nature, however, remote sensors often produce noisy signals. Several of the remote measurement technologies will be discussed here. It will be pointed out that the remote sensors have not been adequately validated to confirm their ability to robustly and accurately measure physiological signals in the presence of noise.

Many of the remote sensors attempt to measure heart rate (HR), heart rate variability (HRV), and/or respiration rate (RR). HR and HRV are the physiological variables traditionally measured by the gold standard electrocardiogram (ECG), a sensor that requires contact with the subject's skin. The pulse oximeter, another contact sensor, is also often used as a gold standard method of measuring HR. RR is traditionally measured by the gold standard nasal cannula, yet another contact sensor. If remote sensors are to be of value, it must be demonstrated, as a first step, that remote sensors can measure physiological variables with accuracy and precision, as compared to gold standard contact sensors.

Gold Standard Electrocardiogram: The ECG measures the difference in voltage between two locations on the subject's body. The change in voltage produced by each contraction of the heart results in sharp deflections in the measured signal[14]. The top trace in Figure 2 shows an ECG signal measured from a subject at rest. Deflections caused by an individual heartbeat are circled in blue. Three deflections occur for each heartbeat: (1) the P-wave, indicating electrical depolarization of the atria (the two small, upper chambers of the heart); (2) the QRS-complex, indicating depolarization

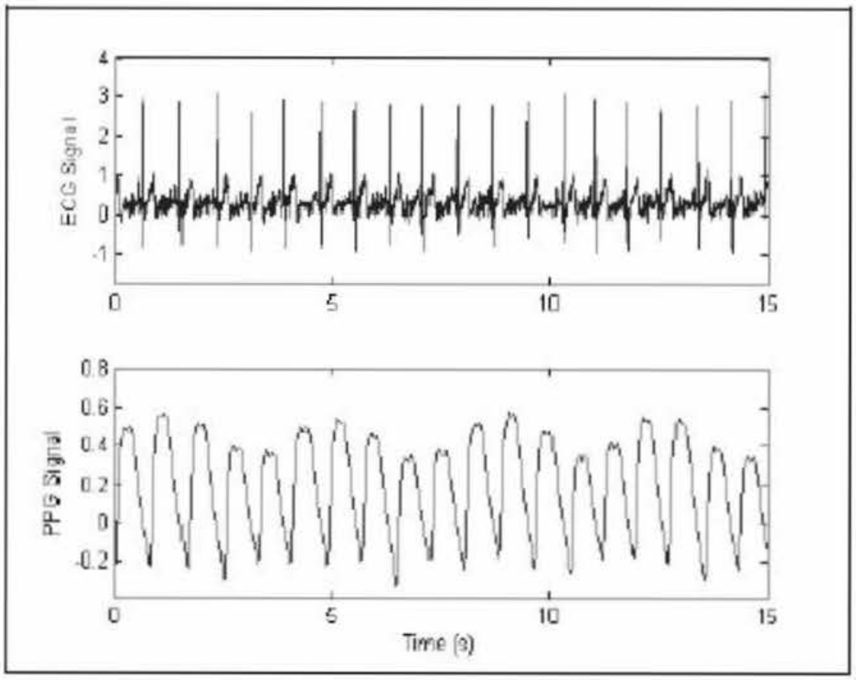

Figure 2: Top Trace: An ECG signal measured from a subject at rest. Deflections caused by an individual heartbeat are circled in blue. P-waves indicate electrical depolarization of the atria while QRS-complexes and T-waves indicate ventricular depolarization and repolarization, respectively. Bottom trace: A pulse oximetry signal measured simultaneously from the same subject. Periodicities in the pulse oximetry signal correlate with periodicities in the ECG signal. Taken from [15].

of the ventricles (the two large, lower chambers of the heart); and (3) the T-wave, indicating electrical repolarization of the ventricles. Characteristics of the P-waves, QRS-complexes, and T-waves have been carefully correlated to different types of electrical disturbances in the heart [14]. For this reason, the ECG is generally accepted to be the gold standard method for assessing electrical cardiac activity. This activity is often summarized in terms of HR and HRV.

Heart Rate: HR is the rate at which the heart beats. HR is traditionally measured as the inverse of the average of the NN intervals, the time intervals between successive, normal QRS-complexes [16]. NN intervals must be measured with respect to a precise fiducial point on each QRS-complex. Techniques have been developed for the automated identification of these fiducial points[17]. As the QRS-complex is a very large, sharp deflection, robust identification of its fiducial point can often be achieved even in the presence of some noise, such as the noise exhibited by the signal in the top trace of Figure 2. However, extreme noise caused from body movements or environmental electromagnetic interference can degrade the signal-to-noise ratio (SNR) of the ECG signal such that HR cannot be measured accurately. For this reason, patients are often advised to remain still when their ECG signals are recorded in the medical clinic. Normal, healthy adults at rest often exhibit heart rates of approximately 70 bpm. During periods of stress, the heart rate can rise well above 100 bpm. Well-trained athletes at rest often exhibit heart rates lower than 50 bpm [11]. Therefore, classifying a subject's HR as "elevated" depends upon knowledge of the subject's baseline HR at rest. For example, while a HR of 72 bpm could indicate a lack of stress in a normal, healthy adult, this same heart rate could indeed indicate stress in a well-trained athlete.

Heart Rate Variability: HRV is a measure of the *variability* between beat-to-beat time intervals and is an indication of the state of the autonomic nervous system. HRV can be measured using both time- and frequencydomain techniques [16]. Time-domain techniques for estimating HRV require robust measurement of successive NN intervals. Therefore, these techniques require robust identification of QRS-complex fiducial points.[7] Some timedomain techniques measure both the high and low frequency components of the NN intervals. Other time-domain techniques measure either the high or low frequency components only. Each technique requires a different mode of interpretation, as the low and high frequency components reveal different information about the state of the subject's autonomic nervous system [16]. Frequency-domain techniques lead to other estimates of HRV, as well. With these techniques, the power spectrum of the NN intervals are estimated, the low and high frequency components are identified in the spectrum, and their relative power is estimated. The ratio of the power in the low frequency component versus the power in the high frequency component is referred to as "the LF/HF ratio" and requires yet another interpretation of autonomic balance [16].

Pulse Oximetry: A pulse oximeter (oxygen-content measurement) shines both red and infrared light through a subject's fingertip, toe, or earlobe. The concentration of oxygen in the subject's blood can then be estimated based upon the difference in absorption between these two frequencies of light [**?**]. Thus the pulse oximeter has become the gold standard method for estimating oxygen concentration [15]. Furthermore, as is shown in Figure 2, the measured signal changes with each heart beat. The bottom trace shows a pulse oximeter signal measured from a subject at rest. The top trace shows an ECG signal measured simultaneously from the same subject. Periodicities in the pulse oximetry signal can be clearly seen. These periodicities correlate visually with the QRS-complexes in the ECG. Thus the pulse oximeter has also become another gold standard method of estimating HR [15]. Provided that the pulse oximeter samples data at a fast enough rate, it is likely that the periodicities in the measured signal could be used to estimate HRV as well as HR. However, little has been published in the peer-reviewed literature describing HRV estimation from pulse oximetry signals.

Nasal Cannula: The nasal cannula consists of small tubes inserted into each nostril. The tubes are used to deliver oxygen to the subject with each inward breath and/or to measure the gases exhaled from the subject with each outward breath. The measured signal is then processed to estimate the subject's RR. Provided that the subject does not breathe through his or her mouth, the nasal cannula provides a direct measurement of the subjects RR [20].

5.2 Remote Technologies for HR, HRV, and RR

The JASON reviewed the following remote technologies for HR, HRV, and RR. These technologies, while representative, do not represent the only remote technologies being developed to measure physiological responses.

Laser Doppler Vibrometry: Several groups have proposed the use of Laser Doppler Vibrometry (LDV) to measure HR and HRV [12, 13, 21]. A laser Doppler vibrometer measures the Doppler shift of a laser light reflected

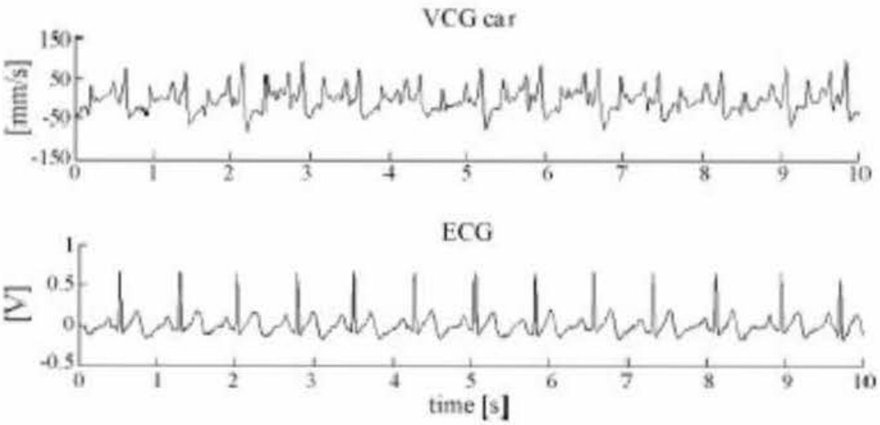

Figure 3: VCG (top) and ECG (bottom) measured simultaneously from the same subject. Periodicities in the VCG correlate visually with periodicities in the ECG. Taken from [13].

from a moving structure. The Doppler shift is related to the component of the velocity of the moving structure in the direction of the laser beam. Thus the velocity of the moving structure can be estimated from the measured Doppler shift [12].

LDV has been used to measure the velocity of the chest wall (directly over the heart) or the skin of the neck (directly over the carotid artery) at a distance of 1-2 m [12, 13]. The resulting signal is referred to as the vibrocardiogram (VCG). Figure 3 shows a 10-second recording of the VCG and ECG measured simultaneously from the same subject. The subject lay still while the data were recorded. The laser light was directed towards the subject's neck, directly over the carotid artery, where a reflector strip was placed. P-waves, QRS-complexes, and T-waves can be identified visually in the ECG. Periodicities in the VCG occur at approximately the same rate as the QRS-complexes in the ECG.

Morbiducci et al. [12] have briefly described in the engineering literature their methods for automated estimation of HR and HRV from the

VCG. They use techniques similar to what are used for the ECG: First, fiducial points are identified in the VCG indicating individual heartbeats. Then, NN intervals are measured as the time intervals between successive fiducial points. Next, HR is estimated as the average of the measured NN intervals. Finally, both time- and frequency-domain techniques are used to calculate different estimates of HRV.

This group has also described in the literature their experiments to validate the VCG-derived estimates of HR and HRV against the gold standard ECG [12, 13]. Reflector strips were placed on the subject's skin, and the subjects lay still while data was recorded. Results showed that the VCGderived estimates of HR and HRV were similar to ECG-derived estimates. However, these results are suspect due to the manner in which the fiducial points were identified in the VCG: A fiducial point was defined as the first local maximum in the VCG occurring *immediately after the fiducial point in the ECG*, the very gold standard against which the VCG was meant to be validated.

To summarize, LDV is a remote sensor that has been used to measure HR and HRV. LDV, if it could be validated, has many advantages to the gold standard ECG for the monitoring of cardiac activity. First, LDV is a non-contact sensor that can measure signals at a short distance (1-2 m), allowing covert measurement of cardiac activity without the subjects knowledge. Also, as LDV is an optical instrument, it is not affected by environmental electromagnetic interference [12, 13]. However, LDV suffers from many disadvantages, many of which make validation of the sensor very challenging. First, the VCG is an extremely noisy signal. Skin is a poor reflector of light and operation of the LDV has not been validated in the absence of reflector strips placed on the skin. LDV cannot be used in situations where the subject is moving, as the Doppler return from gross body movements is much larger than the Doppler return of the cardiac contractions. Such clutter results in complications when identifying fiducial points in the measured signal, the first step in estimating HR and HRV.

The VCG has not been validated to an acceptable standard. Although two validation experiments have been reported in peer-reviewed engineering journals, both experiments were flawed due to the biased nature in which fiducial points were identified in the VCG. Also, only a very small number of subjects were included in each experiment, leading to statistical analyses that were likely underpowered.

Radar: Radar is another remote sensor that has been used to monitor physiological signals [18, 22]. A radar system emits radio waves or microwaves and then measures the scatter of the reflected waves in order to gain information about the reflective target.

RadarFlash is a radar-based system used to detect the respiration of a person hidden behind an opaque wall [18]. RadarFlash is intended for applications in law enforcement and search-and-rescue. Filters were designed to attenuate frequency components outside the bandwidth of the expected respiration signal, including components caused by radar self motion and fluorescent lights. Figure 4 shows a 55 second recording of the filtered signal reflected through a 20 cm hollow core concrete wall. The system began recording data before the subject walked into the radar beam. Point A describes the filtered signal reflected from an empty room, in the absence of a subject. The subject entered the beam at the point labeled B, near second eight, and then left the beam near second 52. The signal level returned to background at point H. Points C, D, E, and F are downward deflections likely caused by the movement of the subject's chest wall during respiration. The subject had been instructed to take a breath once approximately every five seconds, similar to the rate at which the downward deflections occurred.

The RadarFlash technology has also been applied to the measurement of HR. An additional filter was designed to separate the frequency components due to respiration from those due to cardiac contractions. The results of these efforts have only briefly been published in peer-reviewed journals. Conference proceedings only briefly describe the use of this technology to measure HR at a distance of 10 m and no validation data were presented [18, 23, 24].

To summarize, radar has been used for remote measurement of both RR and HR. If it could be validated, the ability to measure these signals remotely, without the knowledge of the subject, is an advantage of radar over the gold standard methods. However, radar also suffers from the same disadvantages as LDV. First, the radar return is an extremely noisy signal. Special care

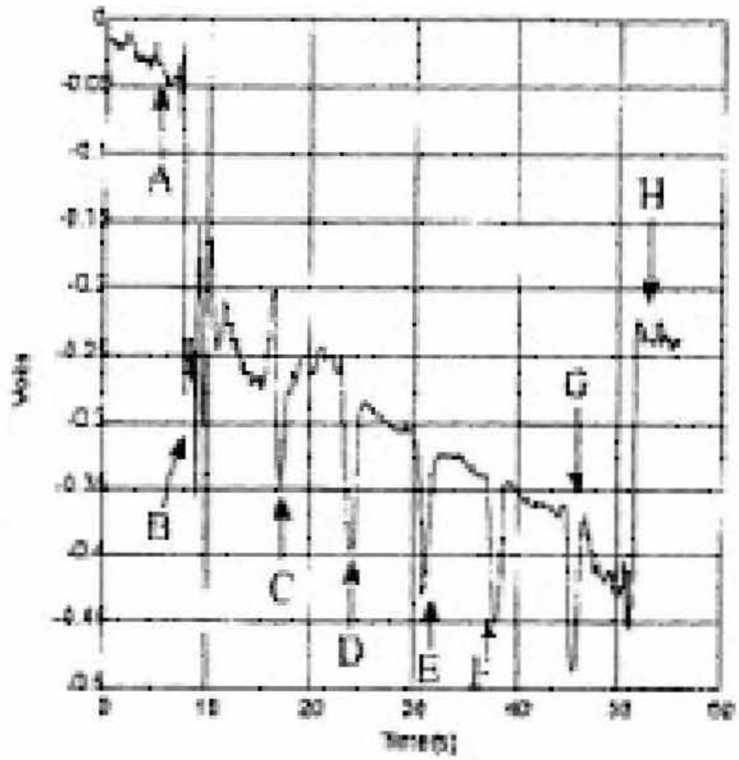

Figure 4: The radar signature of one subject's respiration pattern, measured through a 20 cm hollow core concrete wall. The subject walked into the radar beam near second eight and walked away from the beam near second 52. The downward deflections marked C, D, E, and F were likely caused by the movement of the subjects chest wall during respiration. Taken from [18]

must be taken to design filters to attenuate the frequency components of gross body movements, as well as radar self motion and other environmental clutter, in order to increase the SNR of the cardiovascular and respiratory information in the reflected signal. Furthermore, no validation experiments have been reported in the peer-reviewed literature.

Thermal Imagery: Another group has proposed the use of thermal imagery to estimate HR [19]. Their system is based upon the premise that blood pulsing through vessels modulates the temperature of surrounding tissue, with greater modulation in the vicinity of major superficial blood vessels, such as the carotid artery.

Garbey [19] describe in the engineering literature their methods for esti-

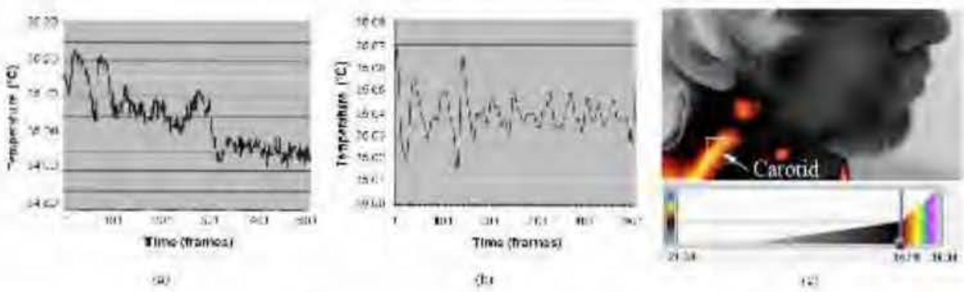

Figure 5: (a) The raw temperature signal recorded from an area directly over the carotid artery, shown in (c). (b) The filtered temperature signal. Taken from [19].

mating HR from thermal images. Thermal video is taken of a subject's face and neck while the subject is still. Tracking software has been developed to allow the thermal video to remain focused on the region of the subject's neck directly over the carotid artery. Once the thermal video is recorded and saved to disk, an analyst studies the first frame of the video and manually draws a line directly over the subject's carotid artery. The pixels recorded in the vicinity of this line are retained for further analysis. Filters have been designed to attenuate the frequency components of the measured signal outside the expected bandwidth of the subject's HR, based on physiological limits. Figure 5(a) shows the raw signal recorded from an area directly over the carotid artery, shown in Figure 5(c). Figure 5(b) shows the filtered signal.

Frequency-domain techniques are then used to estimate the HR from the filtered temperature signals [19]. For each video frame, Fast Fourier Transforms are applied to pixels along the length of the carotid artery and the average over all Fast Fourier Transforms is calculated. The HR is identified as that frequency with the highest energy within a pre-determined bandwidth, based on physiological limits. As all processing is accomplished in the frequency domain, identification of individual fiducial points in the time domain is not required, side-stepping one of the largest challenges faced by other approaches [12, 13, 18].

Garbey [19] published the results of their validation experiment. Thermal video and pulse oximetry were recorded for each of 34 subjects at rest.

The thermal video-derived estimate of HR showed a high degree of correlation with the pulse oximetry-derived estimate. Statistical tests were not performed, however.

To summarize, frequency-domain signal processing techniques have been used to estimate HR from thermal images and have shown a high degree of similarity to gold standard estimates of HR. If it could be validated, the ability of thermal imagery to estimate HR without direct contact with the subject is an advantage over gold standard methods. However, the thermal camera must be placed rather close to the subject, at a distance of no more than a few feet; this may lower its utility in some operational scenarios. Furthermore, thermal imagery suffers from the disadvantage of being an extremely noisy signal. Tracking software must be carefully designed, as "the noise introduced due to tracking imperfections causes damage to the performance of the method that is comparable (if not larger) than that of an unclear vessel imprint."[19] Filters must also be carefully designed to attenuate frequencies caused by environmental clutter.

In conclusion, the ability of remote sensors to estimate HR, HRV, and RR must be validated against the gold standard methods. Accurate and precise estimation of these physiological variables using remote sensors will prove challenging. Nevertheless, efforts should be made to validate in a careful manner the physiological variables estimated from remotely measured signals and to report validation results in peer-reviewed medical journals. Efforts should be made to increase the SNR of the physiological information contained in remotely measured signals. Improved software for tracking and filtering may be one option. Efforts should also be made to estimate the HR of a subject from a remotely measured signal using signal processing methods that do *not* depend upon the identification of fiducial points in individual heart beats, as the low SNR of remotely measured signals complicates the robust identification of individual fiducial points. Similarly, efforts should be made to estimate RR using methods that do *not* depend upon identification of fiducial points in individual breaths. Frequency-domain techniques may be one option for accomplishing this goal. Such efforts may result in low-fidelity estimates of HR and RR. However, an estimate of HR rounded to the nearest 10 bpm, or an RR estimate rounded to the nearest 10 breaths per minute, may still be useful for the remote detection of hostile intent. Finally, efforts should be abandoned to estimate HRV from remotely measured signals, as the SNR of such signals is likely to be too low to estimate HRV in an operational scenario. Furthermore, the many different techniques used to estimate HRV require different interpretations that may unnecessarily complicate the soldier's mission.

5.3 Thermal Imaging for Fear and Deception Detection

In addition to measuring heart rate, thermal imaging to monitor changes in skin temperature has been proposed as a means of monitoring stress in individuals, thereby potentially providing evidence of fear or deception. The idea is that key parts of the body, such as the region around the eyes (periorbital region), exhibit anomalous changes in skin temperature due to changes in blood flow induced by stress.

Tsiamyrtzis [42], is among the most quantitative studies of the use of thermography to obtain indications of deception. The authors inferred a success rate of 87.2% in correctly classifying instances of deceptive and nondeceptive behavior. The authors acknowledge that their experiments are somewhat artificial; but, to their credit, they have paid attention to ensuring that each subject is operating under the assumption that there is both a reward for success (either deceivers or non-deceivers being considered nondeceptive) and a penalty for failure (either deceivers or non-deceivers being deemed deceptive).

This study is based on prior work proposing that changes in periorbital temperature can be an indication of fear [32], stress or deception [36, 39]. The underlying idea, that changes in blood flow can provide indications of stress and therefore deceit, seems reasonable, in that functional magnetic resonance imaging (fMRI) and other techniques provide indications of stress-correlated changes in patterns of synaptic firing and concomitant changes in blood flow [31, 29, 28].

Specifically, Pavlidis and colleagues have been developing methods for tracking HR and RR, as well as skin-temperature changes, by means of thermography: that is, by using an infrared-sensing video camera to monitor an individual subject [34, 41, 26, 19, 35]. In a slight modification of the thermal method, they have also used an IR camera to monitor breathing rate by way of exhaled CO_2 [27].

Pavlidis' group compared the results of thermography against those of current technologies, such as galvanic skin response, pressure sensors, and thermistors. They note that one of the key benefits to thermography is that – unlike the polygraph that requires physically attaching the subject (e.g., with wires and cuffs) to monitoring equipment – IR imaging provides a remote monitoring method that is far less intrusive (and uncomfortable) and therefore potentially more reliable than the polygraph [38, 26, 19]. In comparison with breathing rate (RR) and skin temperature, heart beats (HR and HRV) may be more difficult to validate in an operational setting (see Section 5.2 above).

One of this group's specific contributions has been in developing algorithms for tracking the infrared image of specific regions of a subject's skin. Their approach, based on tracking more than one part of the image (e.g., tracking at low resolution for robustness, complemented by tracking at high resolution for reliability), has been successful in a laboratory setting and appears promising for field deployment [34, 42, 25, 26].

The evidence that stress, let alone deceit or fear, can be revealed through thermal imaging is poorly documented in the current literature. Few of the papers are in the peer-reviewed literature, and almost none of them provide enough information to allow independent replication of the experiments or even to allow the measurements to be reanalyzed. For example, Pollina [39], do not provide any data, but simply present aggregated results (cf. their Figures 2 and 3, and Table 1). Assuming that error bars shown correspond to one standard deviation, it is hard to discern any basis for concluding differences between deceptive and non-deceptive group responses, contrary to what is stated in the text. In addition, the magnitudes of the averaged signals being compared are far smaller than the resolution of the instrument used, and there is no comment about the sign of the effect being apparently the opposite of what is expected: a diminished skin surface temperature change for the deceptive group responding to relevant questions (Figure 3 of [39, 36]).

The publications appearing in the highest-profile peer-reviewed journals, [32, 36], do not contain any data. Therefore, it is impossible to evaluate the significance or reliability of the observations, analysis or final conclusions. Indeed, Pavlidis [37] all but retracted the implied significance of their paper published a month earlier, and it is only through an independent commentary that one learns about serious shortcomings of the way in which the study was conducted [29]. Nevertheless, Pavlidis, [36] is presented in subsequent publications as describing objective testing that validates the use of thermography for revealing stress if not deceit [42]; that there is no mention of the retraction [38] is only one of many troubling indications of unreliable documentation.

Tsiamyrtzis [42] is the one publication that provides a summary of data, and here one finds additional indications of faulty methodology. Taking the results in Table 1 at face value, the overall success rate for identifying deceptive and non-deceptive subjects is 54% and 60%: barely above the threshold of random chance (50%), especially if one considers plausible uncertainties (a change in signal for only one subject corresponds to ± 4% and ± 7% relative to the values given). It would thus be more meaningful to score the results relative to random chance, the latter being assigned a zero score (e.g., +100% would mean perfect prediction, and -100% would mean always-incorrect predictions: on this scale, the overall success rate stated as 56.4% would instead lie at 13 ±5%).

The high success rate ultimately claimed by Tsiamyrtzis [42] appears only after applying a low-pass filter to the temporal data. Not enough information is provided to allow independent analysis of this filtering, but it appears from their Figure 17 that the highest frequencies retained in the signal lie in the 1-10 Hz range. There is no discussion of why such an averaging is needed, or why it caused so many of the predictions to change (14 out of 39, of which 10 were for deception and 4 for non-deception, corresponding to correcting 9 false negatives and 4 false positives and creating 1 false negative).

Aside from these methodological concerns, the evidence that deception causes increased blood flow observable by thermography is weak. Initial studies using fMRI and other methods have shown only partial correlation between deception and measured brain activity for groups of subjects [31],[33], and follow-on work has revealed individual responses to be poorly predicted [30]. Of course, correlation is not a proof of causation, and it is troubling that there is so little evidence of the strength (and perhaps even the sign) of the measured signal being proportional to the degree of stress or deception. Work on other animals is not reassuring: though more intrusive measures (and inducements) of stress are possible, there is still a considerable gap with respect to validating thermography [40]. Moreover, Haynes and Rees [28] point out that current techniques for imaging brain activity (such as fMRI and PET) do not have sufficient spatial resolution to track the actual neurological processes thought to be induced by stress or deception, and it is even less likely that the spatially-averaged signal provided by thermography retains as much relevant information as the brain imaging.

In summary, the skepticism expressed in the National Research Council [3] report, about thermography as well as other technologies being applied to screen individuals for signs of deception, remains. The past half-dozen years of publications have not advanced the scientific validation of thermography to any significant degree.

5.4 Other Sensors

A variety of other sensors for monitoring physiological responses and behavioral responses were presented to the JASON. All the sensor development reviewed suffers from lack of scientific rigor in the evaluation of the sensors to measure a response and to correlate the response to a behavior of interest. In all cases, the technology is far from fully developed. In some cases, the technology is already deployed and the JASON questions the utility. These are summarized in this section.

Eye tracking analysis. The work of Alfred L. Yarbus in the 1950s and 1960s [43] established that tracking the focal point of the eyes gives an indication of the task given to a subject, and that tracking the gaze of a subject's eyes can reveal their interests in an image. Normal eye movements do not progress from one visual feature to another with a smooth and even speed. Rather, the eyes fixate on a location for 200 to 350 milliseconds (fixations), then rapidly relocate to a new location (saccades). One might imagine that changes in the timings of fixations and saccades when a subject is under stress, and/or changes in the pattern of fixations and saccades as a subject's gaze explores an image could be informative.

Early studies on eye tracking made use of a form of contact lens with pointers extending from the lens to reveal gaze locations. Researchers studying sleep pattern in humans make use of "electrooculogram" devices that record eye movement (but obviously not gaze direction) by evaluating the electrical stimuli that trigger movement of eye muscles, which rapidly operate, for example, during REM sleep. These more invasive eye-tracking technologies are either too inaccurate or are impractical for applications requiring rapid and high throughput evaluation of subjects.

More recently, non-contact systems for eye tracking have been developed. For example, infrared (IR) and near infrared (NIR) light is used to create a corneal reflection (CR). Also, features of the eye (such as retinal blood vessels) can be used to track eye movement. Both methods can be used to reveal the subject's gaze location in a stand-off fashion using video cameras. However, a brief calibration period is required, which in most instances will reveal to a person that their eyes will be subjected to tracking. Eye tracking systems that use IR are less effective outdoors in ambient light due to natural IR radiation. Also, data collected on subjects whose head is free to move during the examination will require correction for head motions if researchers wish for the eye tracking results to accurately predict the location of the subject's gaze.

A common application of eye tracking recently is in the evaluation of web page design. Web page developers evaluate user search patterns as they explore a web site, thus allowing optimal web page design. In addition, companies such as Eye Tracking, Inc. claim to be able to equate pupil diameter with mental effort. If true, this would allow researchers to assess the level of concentration a subject is devoting to a particular visual feature.

Despite these apparent useful applications of stand-off eye tracking technologies, there is no scientific validation that eye tracking can yield high confidence predictions of deception or intent. Even if validation of eye tracking for deception or intent were achieved, protocols would need to be developed for evaluating subjects in real world settings.

Visual recording and analysis Visual cues are likely to hold some value in assessing the intent to do harm, or at least a stressed state. These range from the more obvious (hand gestures or "emblems", agitated or atypical body movements) to the more subtle (facial micro expressions and eye movements). One of the more developed areas of research is in human facial micro expressions that claims to give cues to the feelings of a person. This area was pioneered by Paul Ekman. In 1978, he published on "FACS" (Facial Action Coding System), which is a system intended to document all possible micro expressions [44].

Ekman currently operates a company (Paul Ekman Group http://www.ekmangrouptraining.com/) that markets instructional tools and training workshops for those interested in using micro expression information for emotion and lie detection. The METT2 (Micro Expression Training Tool) has shown some utility in improving the ability of trainees to increase their accuracy in micro expression interpretation. The training phase gave useful visual and voice instructions as to the meanings of various facial expressions. However, scoring of the trainee's abilities to assess micro expressions is conducted by briefly flashing the image of a particular extreme facial expression to interrupt an image of a normal face. This training tool seems to test for a trainee's ability to focus intently on a face for an unrealistically short flash of a maximally formed facial expression. In this regard, the term "micro expression" seems to be a misnomer, as the expressions used for training and testing by METT2 are not small or subtle. This training system actually is better for training individuals to make rapid assessments of fully-formed facial expressions.

There is some evidence [45] that some of the micro expressions appear to have the same meaning in different cultural backgrounds. However, it is not clear whether the key facial expressions that are the most indicative of hostility, such as anger and disdain, are universal [46] and could be used in any cultural background to help predict a subject's emotional state. However, no evidence exists that even this knowledge alone can yield an accurate prediction of intent.

Audio recording and analysis. There is a long history of claims of successful lie detection made by representatives of companies who sell voice stress analysis (VSA) systems. Unfortunately, the few scientific tests of this technology [47] - [49]) have revealed that these systems are essentially useless for making accurate predictions of deception. However, there might be some psychological effect, termed a "bogus pipeline" effect, wherein subjects under examination might be more truthful if they believe that a lie detection device actually works [47]. This appears to be similar to the "placebo effect" in medicine, where patients given a pill lacking a pharmaceutical agent state that they feel better, believing that they are receiving pills with an active ingredient.

In a recent study of VSA [47], jailed subjects were examined while addressing questions on their recent use of five illegal drugs. This study is particularly powerful because the researchers have a high probability of establishing ground truth regarding the subject's drug use because they administer drug tests as part of the study. Although the drug tests can give a false positive or false negative result, the accuracy of these tests is believed to be near 100%. Two types of analytical systems used were based on layered voice analysis (LVA) and computerized voice stress analysis (CVSA). In a study of 319 individuals, only 20% and 8% of the deceptive subjects were identified by LVA and CVSA, respectively. Furthermore, the ratio of false positive to true positive predictions was 5.0 and 14.1, for LVA and CVSA, respectively. Thus, these systems are remarkably inaccurate in their predictions, with between 80% and 90% of predictions of deception being incorrect.

Marketers of LVA systems point out that LVA is distinct from typical VSA because LVA examines the entire spectrum of sounds created during vocalization. Sales literature from marketers of LVA systems state that emotions can be detected by analyzing vocalization sounds because brain activity leaves traces in voice patterns. Therefore, they argue that stress and a multitude of other emotions can be detected in voice patterns. However, scientific support for these claims is lacking. In contrast, VSA purports to analyze the "infrasonic" or "micro tremor" component of the human voice, which supposedly yields vibrations between 14 and 8 hertz. Changes in vibrations in this range have been proposed to occur by involuntary tensing of the laryngeal muscles during stress. Again, no validation of this mechanism or the utility of measuring sounds in this range to detect deception have been published in the scientific literature, despite much effort over two decades [50]. This appears to be true for CVSA as well as its non-computerized precursor called PSE (Psychological Stress Evaluator).

5.5 Future Attribute Screening Technologies Program

Integration of multiple sensors of cues that are indicative of the intention or desire to cause harm, particularly those with stand-off and rapid output capabilities, could improve the accuracy of such predictions. This is the goal of the "FAST" (Future Attribute Screening Technologies) program.

The FAST program is sponsored and coordinated by HSARPA and is being assembled at Draper Laboratory. The unit or "Box" is being assembled as a portable screening system that integrates typical screening systems (e.g. a metal detector and two human screeners) with additional stand-off devices that monitor in seven areas: 1) heart rate; 2) respiration rate; 3) eye tracking; 4) electrodermal activity (open pore count); 5) thermography (facial temperature changes; 6) video recording (no firm plan for what will be analyzed); 7) audio recording (voice pitch to assess stress).

A long-standing criticismof lie-detectordevicesis that they lack a theory to rationalize why the cues monitored by these machines (heart rate, respiration and electrodermal activity) should be indicative of deception (National Research Council [3]). The HSARPA program is attempting to resolve this for the devices they intend to integrate into the FAST Box by formulating a theory for "malintent", or the intent to do harm. The developers have conducted a literature review[5] to identify cues that are indicative of malintent. Stand-off sensor devices that possibly can rapidly evaluate these cues were then selected for integration into the system.

One attractive aspect of the FAST Box, as currently envisioned, is that it could incorporate any number of validated technologies that evaluate cues indicative of intent. Each additional technology would supplement screening technologies that have proven to be successful (now restricted to metal detectors and human screeners). New technologies grafted to the FAST Box as it evolves over time can then be field tested with comparison to technologies that are already being used. Unfortunately, the seven sensor

[5] The meta analysis was unpublished at the time of this report.

technology areas mentioned above have little or no validation for predicting malintent. Furthermore, at least one technology (voice stress analysis) has no credible scientific support for utility in lie detection even after more than 30 years of use. Problems with several systems being grafted into the FAST Box are summarized below.

Eye Tracking Analysis. Currently, the FAST Box developers are using a Tobii eye tracker, produced by Tobii Technologies AB in Sweden.

This system uses CR technology, but requires user intervention to calibrate the system. The FAST Box developers plan to eventuallyuse a product called "Smart Eye Pro 4.0", that is also produced by a company in Sweden called Smart Eye AB. The system uses CR technology and records the subject using six cameras. Head motion is compensated by generation of a 3D head model (involving identification of landmark features by an operator). The company claims to use "an innovative IR lighting approach, which suppresses the effect of sunlight and shadows, and enables the system to work in both daylight and in complete darkness without recalibration". However, calibration for each person being screened appears to be necessary. In addition, it is not clear how the FAST Box developers will use this eye tracking technology to observe signs of malintent.

Micro Expression Analysis. Regardless of the pros and cons of the METT2 training system, the FAST Box could make use of FACS to help evaluate facial expressions of emotion. At the time of this report was written, the developers of the FAST Box had not resolved how micro expression analysis will be conducted. Presumably, the FAST Box would either need to have very attentive human operators carefully assessing the facial expressions of each screened subject, perhaps aided by replaying the recorded images of faces in slow motion, or the system would need to be automated. The use of trained operators to make these assessments would likely increase the expense of the security system and increase the time needed to screen each individual. Thus a more attractive feature would be to integrate an automated facial analysis system that could rapidly and uniformly assess facial expressions.

Progress is being made in the use of computer algorithms to make assessments of facial expressions [50]-[52]. If the key features of the face can be extracted from a video image, then these feature patterns can be automatically matched according to FACS. For example, feature extraction can be achieved by "difference imaging", or assessing the difference in pixel intensity value from frame to frame of a video recording. For FACS assessment, the "optical flow" of the features revealed by difference imaging can be used to determine the motion of facial features. Establishing the difference image and optical flow of the entire face is computationally demanding, while tracking individual features requires less computational power.

Unfortunately, head movements often can accompany the formation of facial expressions and cause changes in image alignment that can confound assessment of facial features. Head movement can be restricted by clamps, or a head-mounted video camera will inherently retain the video perspective regardless of head movements. However, these remedies would not be practical in a FAST Box that seeks to rapidly screen subjects. Large head movements could be compensated, for example, by modeling the head as a cylinder and estimating all possible degrees of motion (translation: horizontal and vertical; scale:forward and backward; rotation: yaw and pitch). These estimates are then used to warp the image to a common perspective before facial features are evaluated.

Although the number of factors that could erode the reliability of facial expressions are large, advances in research and development should allow systems to be created that could be useful for the FAST Box. One automated system has been recently shown to be predictive of the difficulty subject have in understanding lectures [46]. However, such automated systems might reliably assess the emotions reflected by facial expressions, but these emotions might not be an accurate predictor of the intent of the subject under evaluation. The Human Factors division of DHS is developing an automated micro expression analysis system in a program called "Project Hostile Intent", and the Draper Laboratory team plans to integrate this system into the FAST Box once completed. Laboratories at UCSD (http://www-cse.ucsd.edu/facresearch/facultyprofiles/CottrellG. html)and the Robotics Institute of Carnegie Mellon University(http://www. ri.cmu.edu/projects/project10.html)are also working to develop automated facial analysis systems that could be considered for use. Even if these automated systems become proficient at analyzing micro expressions, the link between emotion analysis and intent prediction has not been validated.

Voice Stress Analysis. Recent efforts have been directed toward the study of multiple vocal parameters to evaluate stress. Even if these systems could reliably detect stress in a voice recording, there is no scientific demonstration that VSA can differentiate between stress due to deception or malintent and other causes of stress. The FAST Box developers are also aware that commercial-off-the-shelf voice analysis systems are not useful for lie detection. Therefore they have begun an internal program at Draper Laboratory to develop an algorithm to assess voice pitch. However, evidence is lacking to prove that voice pitch changes can reliably predict deception or malintent, and it is not clear how the Draper Laboratory program will resolve this problem.

Given that voice analysis systems have not been scientifically validated and show exceedingly poor rates of successful prediction of deceit, it is unlikely that FAST Box architectures of the near future will benefit from installation of these devices. However, the developers of the FAST Box might choose to take advantage of the bogus pipeline effect [49]. Installation of a cardboard box covered in aluminum foil would be far cheaper that the purchase of a VSA system, but still could be used to make subjects believe that a functional lie detector is in place, thereby improving the likelihood that subjects will tell the truth.

6 INTERVIEW METHODOLOGIES

Several interviewing techniques were presented to the JASON. These methods were presented as holding promise for use in deception and intent detection. Extremely limited studies have been conducted to determine the utility of the methods in the context of their current use. However, there is anecdotal evidence that these methods involve excellent police work and may be able to be validated and developed further for deception detection.

Findings:

- Quoted measures of success for detecting deceit and intent are *post-hoc* and incorrectly equates success in indentifying covert combatants with drug smugglers, warrant violators, etc.
- Detection of deception through interrogation is further confounded by culture.

Recommendations:

- Scientifically assess anomaly detection strategies.
- Establish success metrics and criteria.
- Scientifically assess utility of interview methodologies for detecting deception.

The JASON was exposed to three interview/interrogation techniques being used to assist law enforcement. These techniques have not been developed for use in a military context. The three methods are described in the remainder of this Section.

6.1 Informal Interviewing

(b)(6) [redacted] (Table 1) has developed an informal interviewing approach to help identify individuals with potentially malicious intent, distinguishing them from a surrounding crowd of innocent bystanders. His approach is based on the hypothesis that he can distinguish "tactical" responses of malefactors versus "limbic" responses of innocent bystanders. His approach has been instantiated by the Transportation Security Administration in their Behavioral Assessment Screening System (BASS) and their Screening of Passengers by Observation Technique (SPOT) training [http://
www.tsa.gov/whatwedo/layers/index.shtm].

Informal interviewing is proposed as a means of probing a crowd of individuals, perturbing it relative to a baseline of innocent activities, and looking for evidence of stress, if not deception. It is plausibly more useful than passive observation in identifying anomalies in activity, but there is no objective demonstration that it is more effective than random sampling in identifying individuals with malevolent intent. Its success is described entirely through anecdote, and there has been no validation of the method with respect to failed detections, false positives or opportunity costs. In particular, it may be that the most likely to be identified as anomalous are the least professional and therefore least dangerous or important targets within a crowd, such that application of the method might actually reduce the chances of identifying the highest-value targets.

This method of informal interview was developed for law-enforcement, to enhance the capabilities of police officers posted in public places such as airport terminals or train stations. The method follows the steps outlined in Section 1.

1. Establish a baseline of activity for the locale.
2. Identify individuals (anomalies) whose actions distinguish them as outliers from this baseline.
3. Informally interview the identified individuals.

4. Send those individuals showing indications of a "tactical profile" inthe informal interview (monotonically increased stress) to secondary screening or other action.

First, a baseline of normal activity must be established. This baseline is intended to characterize the normal flow of individuals, as well as the range of interactions typically occurring among individuals in a particular locale (or lack of interactions, as might typically be the case in a public venue). Are there individuals who hesitate or change direction when coming in the direction of uniformed officers? Do any individuals exhibit obvious signs of nervousness, of extreme violence or contempt, or of being "dead to the world" (i.e., about to die)? Are there groups of individuals moving together and, if so, are they overt or less-than-obvious as a group and do they show any indication of military training in their activity?

These considerations are highly subjective, and may be prone to considerable false positives. That is acceptable, from (b)(6) perspective, because the main purpose of identifying a baseline is to provide a basis for focusing on a subset of individuals who warrant further consideration (as distinct from the rest of the crowd). A benefit to (b)(6) approach is that he emphasizes the need to explicitly document – no matter how the baseline was established – why an individual or group is identified as an outlier. This imposes a discipline, with the officer identifying specific behavior that is (perceived as being) outside the norm, and mitigates the chances of racial profiling or other unreliable preconceptions being the basis for focusing on specific individuals out of the crowd.

Individuals identified as outliers are then interviewed in an informal manner in order to see whether there is cause for further attention. The initial criterion does not need to rise to the U.S. legal standard of "probable cause" and false positives are acceptable because the interviewing method is completely voluntary. (b)(6) emphasizes that to make this interview useful it must be as friendly as possible, in order to mitigate stresses associated with the environment (e.g., a terminal crowded with people on tight schedules) as well as with the reality of being approached by a police officer (uniformed or not).

Two key criteria appear to be involved with the steps in (b)(6) approach. First, if during the informal interview the individual (or group) tentatively identified as an outlier reveals inconsistent information, this warrants further investigation. Depending on the degree of inconsistency, ranging from a simple mistake that is easily corrected to indications of willful deceit, the officer may simply extend the informal interview or else direct the individual to secondary screening or other follow-on action. This approach of looking for inconsistencies in individuals' responses is part of traditional police work.

The second criterion is based on (b)(6) hypothesis that an innocent bystander has a "limbic" response, whereby stress first increases at the onset of the interview because of the approach by a police officer and then decreases in response to the friendly nature of the interview. The officer's objective is to be viewed as a helpful friend, whether or not the stress level starts high (for example because of the environment), such that an innocent individual would exhibit a reduction in stress. In contrast, a "tactical" profile is one of ever-increasing stress because the individual intends to do harm and the officer represents an obstacle to completing that mission. In fact, a well-trained "tactical profile" may start at a low stress level if the individual is well trained for the mission.

(b)(6) has characterized these responses in terms of changes in heart rate with time, though the profiles are more notional than quantitative (personal communication, June 19, 2008). The key is to identify those showing monotonically increasing evidence of stress as being the individuals worthy of further attention. This is the third component of (b)(6) method and, depending on circumstances, that additional attention may range from secondary screening to denying access to the facility (e.g., airport or train station) to arrest. Only in the last case is there a need to meet the U.S. legal criterion of "probable cause."

This approach involves three levels of sorting in identifying potential malefactors: those who are outliers relative to the background, so experience an informal interview; those who exhibit a "tactical" response to the interview; and those who then fail at secondary screening or other follow-on action. Each level of sorting narrows down the group of individuals to be considered and, because there are 3 distinct levels, it is thought that false positives are acceptable at each step because innocent individuals will ultimately be exonerated.

In summary, the method follows a model of observing anomalies that is common in law enforcement, so is based on considerable experience in realworld applications. That model requires both establishment of a baseline and probing of the group of people (via the interview) in order to validate anomalous behavior. One of its key features is the requirement that the interview be friendly and informal – this is essential in order to distinguish tactical from limbic responses, and is entirely consistent with approaches recommended for counter-insurgency [53, 54]. Therefore, the approach may be useful for military applications responding to law-enforcement and insurgency challenges.

Evidence of success provided was not based on criteria established beforehand, but depended on after-the-fact evaluation that individuals who were identified as anomalous were then found to be guilty of some transgression. There is no evidence that the strength of the signal is proportional to the significance of the transgression. Therefore, even if there is a correlation between anomalies in behavior (relative to a baseline) and some level of guilt, there is no clear indication of a causative link between anomaly and degree of transgression. Finally, it is not clear that the underlying model of limbic versus tactical response is valid; or, if valid, that it is being reliably observed by the interviewing officer.

6.2 Cognitive Interview

A second interview method being explored for use in deception detection is the cognitive interview (CI). The CI was developed in the 1980's [55, 56] based on the evolving theory of cognitive recall. The initial development was for enhancing eyewitness recall. In that context, the method has been both laboratory and field tested [57, 58, 59] and has demonstrated value for eliciting useful and complete information from eyewitnesses for recently committed crimes.

Throughout the last three decades, cognitive theory and CI methodology has been adapted to improve survey questionnaire design [61, 62, 63]. These adaptations have been scientifically studied through well designed statistical experiments in the laboratory and in the field, yielding a range of results regarding the degree of carryover effects between laboratory and field tests [60, 63, 65, 66].

The CI attempts to aid an individual's memory recall through asking the witness to first re-create the environment surrounding the event, including external (e.g. weather), emotions (e.g., personal feelings), and cognition (e.g., relevant thoughts) environments. Next, the witness is asked to do a straightforward memory retrieval of the event. To further help the memory process the witness is given more opportunities to recount the event, frequently in a different chronological order. Finally, the witness is asked to view and recount the event from a different perspective. The questioning is open ended, allowing many opportunities to remember details. Lawyers and law enforcement have found the CI to yield valuable information.

Research is emerging based on the hypothesis that someone engaged in deception has an increased cognitive load [67]. (b)(6) (Table 1) has tried to integrate that hypothesis with the cognitive interview. Given the open-ended nature of the questioning and the amount language (text) generated during the CI process, it is hypothesized that a linguistic analysis could separate deception from truth. Whether or not the CI can be used effectively in this setting has yet to be validated. Fisher [68], one of the founders of the CI, has commented on the fact that translating from the laboratory to the field for the cognitive load hypothesis in the context of deception will be much more difficult than for the CI in the context of eyewitness recounts, where one is only replicating memory recall.

6.3 Technology Assisted Interviewing

Thermal imagery is currently being used in some police interrogation rooms to monitor suspects during questioning. The impetus to use thermal imagery was derived from a metabolic study using this technology at the Mayo Clinic. The study revealed a rapid thermal response to stress caused by a laboratory accident involving broken glassware.

In a police interrogation room in Gwinnett county, Georgia, the camera is concealed behind a false fire alarm panel several feet from the suspect. The camera is a modified commercial camera that can be adjusted to give a readout over a narrow temperature range (9 degrees F) centered on the subject's average temperature, which allows high sensitivity to small changes in temperature. During questioning, a remote operator examines thermal imagery in real time and covertly relays assessments regarding the suspects stress levels, as presumably indicated by visual inspection of the thermal imagery and corresponding data. The procedure allows the questioner to continue the lines of questioning that might be most productive.

The use of the thermal camera has caused the officers to alter their questioning techniques. Most importantly, questioning strategies must be designed to stress guilty subjects when they are being deceptive, but calm innocent subjects when they are truthful. The questioning style follows a process similar to a cognitive interview (CI). Subjects are asked to describe their story several times in increasing detail. The officers believe untruthful subjects experience increased stress due to the difficulty in remembering and elaborating on their lies. Breaks in pursuing a topic causing a physiological stress response are required to allow the suspect to return to a normal temperature state before returning to the stressful line of questioning to confirm sensitivity in to specific topics. Operators call this tactic "catch and release" questioning.

Temperature changes of up to 3 degrees F have been seen, but can be influenced by other physiological complications (sunburn, irritants, age, make-up). Also, it was noted by (b)(6) (Table 1) that two or three "sociopaths" have completely failed to trigger signals when subjected to thermal camera analysis. Stories regarding the utility of this method and other methods used by police interrogators are entertaining and compelling. However, there is a lack of study data to provide support for these claims or to assess the success rates for predictions based on thermal imagery.

State laws in Georgia apply regarding use and storage of the information collected by the Gwinnett County Sheriff's Department, and there indeed seems to be a wealth of data available that could potentially by used to assess the utility of thermal imagery in this context. The data cannot be used as evidence in court by the prosecution, suggesting that the technique is too new and unproven, so that laws have not been crafted to exploit the technology. A company has been founded (Detection Dynamics, Inc.) to exploit thermal imagery for police interrogation use and they are privately funded.

6.4 Cultural Impact on Ability to Detect Deception and Infer Intent

Cultural implicationsof interrogation methods have not been adequately studied (Intelligence Science Board)[4]. In the military setting, an interview to detect deception must be conducted keeping cultural nuances in mind, or it will likely fail. There are examples of volunteered information or an interrogation, in which a person identifies "bad guys", only to be revealed later that the "bad guys" were people that the person interrogated wanted eliminated for his own purposes. If manifestations of stress are used to detect lying, intent, etc., it should be borne in mind that the person being interrogated who has been bred to have the *ability* to deceive from a tender age, will have no trouble identifying the situation as one that calls for this ability. He or she is likely to have little trouble lying/deceiving with little stress, or manifestations of stress, and if this is what the interrogation tools are designed to detect, they will produce no information or false information. In particular, information given, especially under duress, is more likely to be false, if it is perceived at the time as to mitigate harsh treatment, or is it perceived at the time to be a moral imperative, i.e., one not accompanied by guilt. As an example, people of Easter-Mediterranean and Middle-Eastern cultures will be much more likely to respond to a situation in which their cultural traits are appreciated and then perhaps exploited, such as placing them in a non-threatening situation and convincing them that personal or family honor is at stake.

Successful interrogation across cultural boundaries is extremely difficult [4]. While training and the exposition to cultural differences will, of course, always help the interrogator, it is too much to ask that a few weeks immersion will enable cultural characteristics to be understood and cultural boundaries to be crossed successfully. Only exceptional and specifically talented people will be successful, often people with foreign-language skills, preferably more than one, acquired at a young age, or people who have been raised in two, or more, sufficiently different cultures, will be suitably sensitive to the issues.

7

ESTABLISHING A SCIENCE BASE: LEARNING THE TRUTH

Throughout this report, the necessity of using a scientific process to determine the truth of experimental claims, and the usefulness of proposed technologies has been stressed. This is because experience shows it is very hard to learn things reliably about human beings, and people fall victim to various forms of wishful thinking. The strength of a science base (peer review, reproducing results, double blind trials, and the like) is that it makes it harder to fool ourselves into accepting things that are not true.

The big cost of getting it wrong is the opportunity cost. In this case effort spent on introducing ineffective technology or ineffective procedures is effort that has been taken away from alternatives, such as training.

Findings:

- No compelling evidence yet exists to support remote monitoring of physiological signals in an operational scenario.
- No scientific evidence exists to support the detection or inference of future behavior, including intent.

Recommendations:

- Establish a discipline of *science* for the development and deployment of new technologies and methodologies.
- Define goals and validate metricsfor both performance and effectiveness in a realistic operational scenario.
- Use *blinded* protocols in performing case/control studies.

7.1

The Science Base

The development of a science base requires a concise statement of the problem, which would include the operational scenario in the case of inferring intent. The problem needs to be able to be parsed into steps to be taken and evaluated that lead toward solutions to the problem. In the case of the technologies discussed here, the very first steps would include demonstrating that signals exist in the most controlled settings. Once that is established, understanding the strength of the signals in noisy environments, and ultimately in the environment of intended use, would need to be established.

The scientificprocess requires controlled experimentationusing accepted protocols to collect and evaluate information. "Design of experiments" must be used when variation is present in an information-gathering exercise. This variation may or may not be under the full control of the experimenter. These experiments typically include randomized trials. The term observational study is used when full control of the experimentis not possible. Design of experiments, including observational studies, has broad application across all the natural, physical, social sciences, and engineering.

The results need to be properly reported such that studies can be replicated. Finally, altered pathways (next steps) must be validated by peer review and community concurrence. The scientific process is never easy or fast. However, it has centuries of affirmation and is the only process by which one can be assured of not venturing into "pseudo science" [69].

Typically experiments show some correlation between input circumstances and some kind of measurable outcome. The physical sciences frequently do case/control experiments, where the effect is measured on two populations, one in the presence of the stimulus, and the other in its absence. This works well when the two populations really are identical, and when the effect is measured independently of the experimenters expectation of what the results should be. This is possible because in the physical sciences generally it is feasible to establish the same initial conditions and overall experimental conditions in both cases, with the two differing only in the stimulus that is hypothesized to cause the effect. Peer reviewers look hard to make sure the experimental setup is valid, and was followed.

This protocol is not easily adapted to the human subjects, or even in the context of medicine. It is impossible to have identical populations because it is impossible to use the same subjects for both the control group and the treated group, as they will remember the first trial or have been altered by it. Hence various sorts of population-based statistically evaluated studies have to be used. There are many well documented dangers in the interpretation of these studies which can lead experimenters to over-interpret or to misinterpret the results, or to accept results from inherently deficient studies. On medicine we have, on the one hand, "carefully controlled research (such as randomized, controlled trials) involving numerical data has proved more dependable for showing us what works and what does not than has reliance upon expert opinions, experience, hunches, or the teachings of those we revere." [70] On the other hand, the same author observes that it's not always practical, "Countless surgical procedures have come and gone over the past few decades without ever having been subjected to controlled scientific scrutiny." This book is about complementary and alternative medicine, and how difficult it is to evaluate the claims and put this sort of medicine on a firm scientific basis. Detecting deception and intent is surely no easier.

As the quote about surgery indicates, we do not mean to denigrate the professional skills of law enforcement, detectives, or Israeli airport interviews, or any of the professionals upon whose skills society relies to protect itself from destructive forces. New sensors and newly revealed phenomena promise to help, and the question is whether they can be made genuinely helpful, or if the proposed uses can provide perceptible improvementover an existing practice. These are questions that can only be answered with carefully defined studies, although it may not be easy to do so.

Answering these questions is important. There's no point in spending money on sensors unless they do, in fact, improve performance. Nor is there any point in training people to exploit new phenomena, unless that training adds value, and the value is measurable. Further, the cost of introducing ineffectual techniques is to rob resources that could be used to make other parts of the systems more effective.

7.2 The Path to Improvement

So what would the ideal path to improved techniques look like? One would expect promising laboratory studies, replicated confirmation by other investigators, publication in peer-reviewed literature, proposed protocols for using the techniques in practice, and at least two stages of verification that the protocols work, e.g., in the laboratory and in the field.

Well designed studies involving human subjects are of two general types, single-blind and double-blind. These types of studies help to mitigate research outcomes from being influenced by either the placebo effect or the observer bias. Blinded research is an important tool in many fields of research including medicine, psychology, social sciences, and forensics.

In single blind experiments, the individual subjects do not know whether they are subject to the stimulus or not. Such information is withheld to avoid the subjects consciously or unconsciously taking actions that would otherwise skew the results. In a single-blind experiment, however, the experimenters do know the facts behind the trials. Single-blind trials can be effective and robust provided that the experimenters in no way use their knowledge to bias their evaluation of outcomes. This may be difficult with human subjects research, where the experimenter has an expectation of what the outcome should be, and might consciously or unconsciously influence the behavior of subjects. Experiments need to be designed so that even inadvertent experimenter bias doesn't affect the results. Peer reviewers must be given enough information to evaluate the results and replicate them, if needed, to assess the validity of the conclusions.

Double-blind experimental protocols provide a higher standard, in that they are designed to minimize the possibility of bias of either the subjects or the experimenters. In a double-blind protocol, during data collection and data analysis, neither the experimenter nor the subject would know which subjects were control subjects and which got the real treatment. A significant effect would then be manifest as two statistically distinct populations in the outcomes between the two different groups of subjects. It is worth noting that 'statistical significance' is not the same as practical significance.' In other words, statistically significant results are not always 'important' or 'large'. No matter how good the statistics are, a tiny improvement is just a tiny improvement.

The following simple example provides some insight into the necessary attention to detail when setting up scientific studies. Suppose the goal is to determine the accuracy of remote HR sensor in a controlled environment. Before collecting any data, a study protocol would need to be developed.
This would include:

- defining the measurement method and metrics for "ground truth" - use of a gold standard;
- defining metrics to be used for the remote sensing method;
- defining data collection methods the underlying population from which subjects will be randomly selected, and the protocol for simultaneous data collection using multiple sets of technicians;
- defining data processing methods to extract ground truth from the gold standard and for the metrics associated with the remote sensor;
- determining the statistical analysis to be used to assess accuracy.

Once the protocol has been established and reviewed to ensure nothing has been overlooked, the experiment would be conducted and data collected according to protocol. Following the experiment the data from the two measurement methods would be processed and statistically analyzed according to the a-priori defined protocols.

The example above did not contain any blinding. Once it is determined that the accuracy to ground truth of the remote sensor signal exists across a wide human demographic range, additional studies would be developed to evaluate the strength of the signal in noisy environments and in response to stimuli. In these cases, blinded experiments would be introduced. If results consistently demonstrate a useful signal-to-noise ratio, then more human behavior (variability) would be designed into the experiments. For medical studies there are standards to use for randomized trials (available at www.consort-statement.org). Many of their criteria are directly applicable to ways of assessing credibility.

Developing experimentsto determinecorrelation between biometric(physiological) measurements and deception or intent is complicated. Careful attention will need to be given to every detail, from properly blinding and ensuring treatments (e.g., protocol for lying and truth telling, interview method) to the statistical methods for establishing correlation.

Translation from the laboratory and from other controlled studies to the field is a critical step. In this process, it is important to remember several basic things. Laboratory studies are necessary, but not sufficient, to establish real-world utility. Correlation is necessary, but not sufficient to establish causality. Field studies will be quasi observational at best. This presents a risk of selection bias which is the distortion of the evidence or data that can occur as a result (frequently unintentional) of the way that the data are collected. Clutter confounds the measurement of physical observables in the real world. Finally, physical state versus current and future state of mind would need to be established to move in the direction of being able to detect deception, intent to do harm, and the decision to act.

There may be some unique opportunities in the military setting for the development of useful field studies. For example, there is the possibility of training against the stimulus by using select groups, such as red teams, thereby creating a quasi-controlled experiment. There is also the possibility of quasi-controlled data collection with training exercises.

7.3 Some Proposed Full-Scale Trials

After all the laboratory studies, suppose there is a promising new protocol for detecting deceit. How might one test it? For a test to have validity it ought to be based on more than college students and monetary rewards. One might implement the protocol at an international airport to find smugglers (meaning people who lied on their customs declaration). People (or families) would be randomly subjected to the new protocol or the usual screening process, and the fractions of revealed smugglers would be compared. The hypothesis is that these people know they are lying, and that they are a sufficient percentage of travelers that there would be statistical significance. That too is testable, because there will be data on how many people are stopped for being suspicious or actually untruthful on their customs declaration.

After all the laboratory studies, suppose there is a promising new passive technique for detecting hostile intent. How might one test it? Hostile events that occur often enough to establish performance are needed. The data collection should not interfere with the behavior of the subject(s). The goal would be evaluation rather than deterrence. There are approximately 40,000 robberies of convenience stores each year in the United States. The sensor could be deployed widely (and covertly) in convenience stores and used over a period of several months. If the passive technique is automatic, then it's just a matter of statistical evaluation of how well it did. If it requires human intervention, then investigators should analyze the tapes (or whatever information the sensor provides) in a "blinded" fashion, without knowing what happened to each subject.

7.4 Goals and Metrics Matter: A note on rare events

For a screening system, useful measures of performance are the positive predicted proportion and the negative predicted proportion. A "positive" is an instance in which the screening system detected the signal of interest, whereas a "negative" is an instance in which the screening system did *not* detect the signal of interest. The positive predicted proportion is the number of true positives divided by the total number of positives, false and true. The negative predicted proportion is the number of true negatives divided by the total number of negatives. These reflect not only the sensor, but the prevalence of the condition of interest in the whole population. For a screening system, a high negative predictive proportion is desired, so that users can believe negative results. A more modest positive predictive proportion may be acceptable, because the false positives will be weeded out by further screening. For instance, mammography is used to screen for breast cancer in women. It is estimated to detect 75 to 90% of the cancers, and correctly detect the absence of cancer about 90 to 95% of the time. Breast cancer is relatively rare, with about 185,000 new cases diagnosed from tens of millions of people screened each year. The negative predicted proportion is 98%, and that makes it a useful screen, because if a negative is returned, one can be fairly sure (50 to 1) theres no cancer, and no further action would be recommended. The positive predicted proportion is quite modest, about 20%. This is viewed as acceptable, if not as large as desired, because the number of patients subject to secondary screening (typically a biopsy) is manageable and a relatively small fraction of the screened population.

A similar analysis would be pertinent to the task of determining hostile intent. To be useful as a screening test one would need to have a high negative predictive proportion, because of the rare expected incidence in the general population. A screening system only needs a high enough positive predictive proportion to make the required further screening acceptable.

8 CONCLUSIONS

This study provided an assessment of the potential utility and efficacy of monitoring and assessing human behavioral neurophysiology and verbal and nonverbal communication to determine human intent in a military context. The process that would need to be implemented in order to identify covert combatants and ultimately infer intended actions was outlined. A key finding in this report is the need to foster and establish a discipline of science in the development and deployment of potential technologies, including interrogation methods, proposed to be useful in the military setting. Some of the key findings and recommendations are repeated here, along with some first steps that might be considered.

Findings:

- No compelling evidence yet exists to support remote monitoring of physiological signals of deception in an operational scenario.
- Detection of deception is further confounded by culture. • Training exists within the DOD for developing intuition and understanding of the adversaries' culture and environment.
- No scientific evidence exists to support the detection or inference of future behavior, including intent.

Recommendations:

- Train the soldier to be aware of diverse cultural environments and complex human dynamics.
- Look for incremental advances that add value to the well-trained soldier, integrating the end-user throughout the R&D processes to avoid unrealistic expectations.
- Establish a discipline of *science* for the development and deployment of new technologies and methodologies.
- Define goals and validate metricsfor both performance and effectiveness in a realistic operational scenario.
- Use *blinded* protocols in performing case/control studies.

- Employ *unbiased* peer review in evaluating the program plan.
- Leverage research expertise in biomedical sensor development.

First Steps:

- Define operational scenarios and measurable near-term goals for behavior assessment.
- Begin scientific assessment of remote biometric sensing:
  - under controlled laboratory conditions,
  - in noisy, real-world conditions, and
  - demonstrate correlations to behavior – it is unrealistic to correlate with intent.
- Begin scientific assessment of anomaly detection following a "double blind" protocol and demonstrating that methods beat simple random sampling of the crowd.
- Begin scientific assessment of the informal interview techniques

A APPENDIX

This appendix includes reviews of some of what the JASON learned about virtual environments. DOD may find it useful to follow some of these technological advances. As with all DOD technology adoption, rigorous scientific validation and demonstration should be applied.

Head-mounted displays Head-mounted displays (HMD) are near to projecting realistic video, and can be outfitted with a positioning system such that the displayed video changes with head position. HMD systems can vary in sophistication to provide increasing performance characteristics. For example, HMD can be monocular (displaying an image to only one eye), or binocular to permit stereoscopic images to be viewed. Humans have approximately a 180° field of view, but most HMD systems use only a small portion of this potential. Depth perception can be simulated in binocular systems by projecting a portion of the image to both eyes. The user's view can be exclusively video imagery, or can be a video overlay on the users actual view (augmented reality). Examples of HMD systems supplying augmented reality can be found in the cockpits of some military aircraft to display night vision or instrument readings. A rugged HMD for augmented reality produced by Liteye Systems (Colorado) is being used by some ground troops. This system uses newly-developed organic light emitting diode (OLED) technology, which appears poised to improve HMD systems substantially. Therefore, some of the advantages and disadvantages of this system are described in more detail below.

OLED research and development is advancing so rapidly that in July of 2008, Japan announced the formation and funding of a consortium of companies to pursue large scale commercialization of the technology. Companies in this consortium seem well positioned to dominate OLED use for the most common applications for video screens (TVs and electronics devices).OLED systems use far less power than older technologies such as liquid crystal displays (LCD). OLED systems have a response time as low as 0.01 milliseconds, which is more than a 100-fold improvement over LCD systems. OLED films can be stacked to overlay red, green and blue pixels for color displays, which increases the resolution of the display over that of cathode ray tube (CRT) displays. Furthermore, the screens of OLED can be deformed without loss of video, and therefore these screens could be used in devices that must flex or in VR systems that would gain from having a curved video screen.

Although such advances reduce the power requirements and might also reduce the weight of the VR device, these advances do not resolve the demand that high-fidelity VR systems will have for computational power and communications. Furthermore, the compounds currently used for OLED can decay over time, thus limiting the lifetime of the display. However, new polymers being applied in OLED have lifetimes that would last longer than other display technologies.

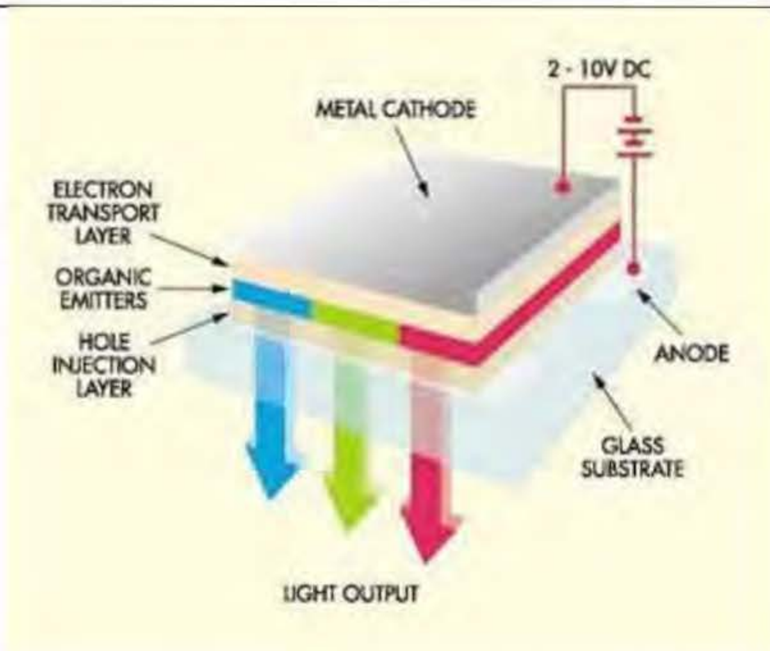

Figure 6: OLED display architecture. Note that the materials can also be made of material that deform without fracture to create displays that bend and retain function.

Audio Devices. Simple audio systems can make use of remote speakers or earphones to deliver monophonic or stereophonic sound. However, more complex sound signatures for a virtual environment could be delivered by a system that uses homophony, or "wave field synthesis" [71]. This technique uses a series of speakers that emit sounds at computer-controlled times to

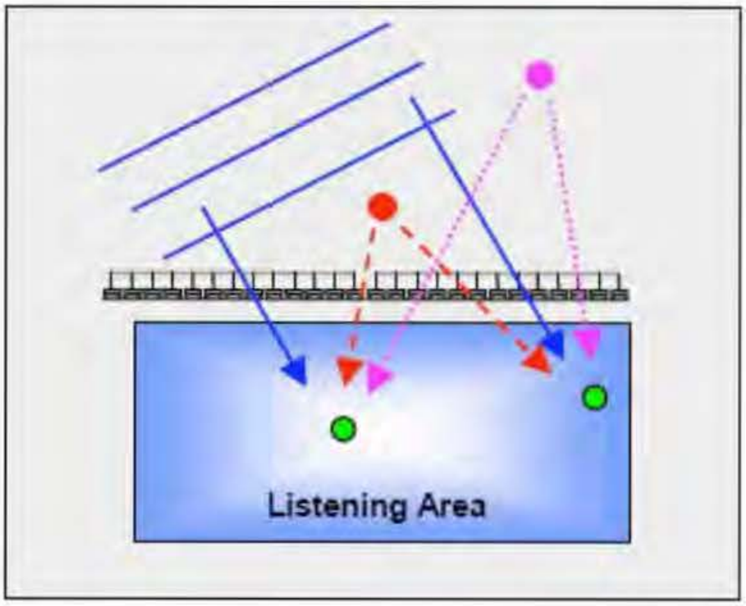

Figure 7: A bank of speakers (black bars) can be controlled to emit sound at specific times to simulate the sound waveform naturally produced at greater distances. In the figure, persons (green dots) in the listening area interpret the synthetic wave field to think that sounds are eminating from various locations (red, pink and blue).

create sound waves that appear to the listener to come from distant points. Sound engineers can use this system to create realistic representations of a large and diverse sound environment to users in a small listening area.

Motion sensor systems. Motion of the head can occur with six degrees of freedom, including forward and reverse motions along X, Y and Z axes, and rotation around these same axes ($\theta_X, \theta_Y, \theta_Z$) (Figure 7). Thus it is a technological challenge to accurately establish the location and the rotation, pitch and yaw of the head. Head tracking is usually achieved by either magnetic or ultrasonic methods [72]. Magnetic sensors function by using a head-mounted receiver that detects magnetic fields emitted from a transmitter. Movement of the receiver in all six degrees of freedom is interpreted relative to the stationary transmitter. However, large metal objects can interfere with the magnetic fields being evaluated, and other factors can introduce positional errors in the predictions made by the system, which need to be corrected [73, 74]. In contrast, ultrasonic systems function by

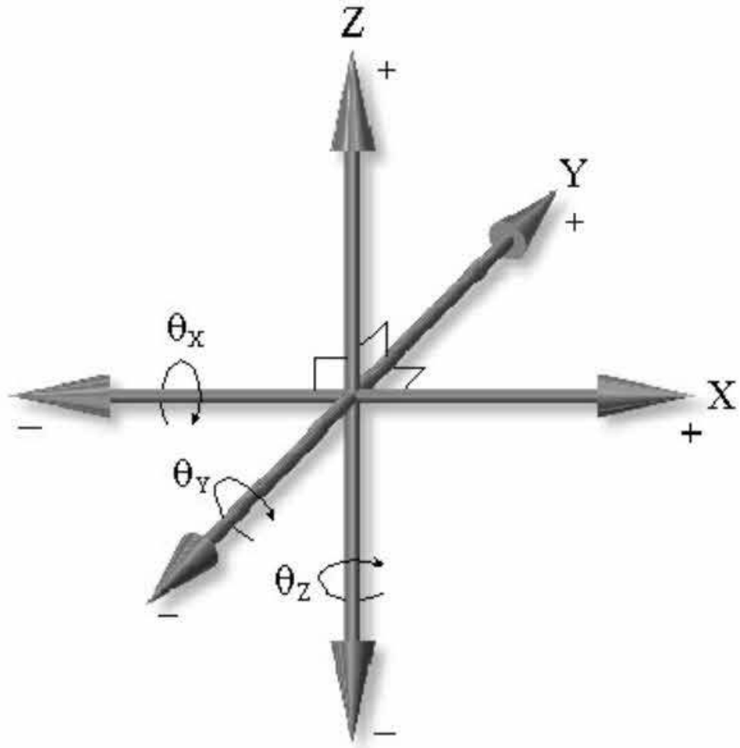

Figure 8: Six degrees of freedom for head motions.

placing three ultrasonic speakers in a triangular pattern around the subject. Reflected sound waves are detected to establish changes in object location.

Data Glove. A "data glove" or "wired glove" is a sensor-laden glove that is worn to track hand movements and can also include outputs to give tactile feedback to the user. Inexpensive data glove systems (e.g. Nintendo's Power Glove) are available for gaming systems that measure finger bending. Bend angles are usually monitored by the changing amount of light that travels through a fiberoptic cable, which decreases with increasing angle of bend. Inertial sensors or magnetic tracking sensors have been integrated with data glove systems to measure hand movement. Specialized data glove systems have been adapted for use in a number of applications including sign language interpretation and remote surgery.

Real-time Video and Audio Input. Some VR applications will require real-time inputs of video and audio. For example, some training scenarios might require a computer to rapidly assess the speech input by the user, as well as hand gestures [redacted] and even facial expressions. The computational time needed to evaluate visual and sound inputs might cause delays in response by the VR system such that the exercise will appear too unrealistic, or will give the trainee an unusually long time to consider subsequent actions. Considerable advances will need to be made for VR systems to seamlessly sense and evaluate all inputs and provide realistic sound, imagery, and other feedback to the user.

References

[1] "U.S. Army Study of the Human Dimension in the Future 2015-1024," TRADOC Pamphlet 525-3-7-01,(2008).

[2] Joyce, G., et al, "Room of Truth," JASON report JSR-03-515 (2003).

[3] *The Polygraph and Lie Detection*. Washington, DC: National Academies Press,(2003).

[4] "Educing Information, Interrogation: Science and Art, Foundations for the Future, Phase 1 Report," Washington, DC: Intelligence Science Board (2006).

[5] Tzu, S. *The Art of War: The Oldest Military Treatise in the World*, translated by Lionel Giles, Harrisburg, PA: Military Service Publishing Company (1944).

[6] "The Role and Status of DoD Red Teaming Activities," Washington, D.C., Defense Science Board (2003).

[7] Meigs, Montgomery M., General, personal communication, (July 2008)

[8] Kortum, P. *HCI Beyond the GUI, Design for Haptic, Speech, Olfactory, and other Nontraditional Interfaces*, edited by Phil Kortum, New York, NY: Morgan Kaufman Publishers (2008).

[9] Committee on Modeling and Simulation for Defense Transformation, National Research Council *Defense Modeling, Simulation, and Analysis Meeting the Challenge*, Washington, DC: National Academies Press (2006).

[10] O'Kane, B., *Badguyology™ Vol 1: Key Factors in Correct Identification of Human Combatant and Non-Combatant*, US Army RDECOM CERDEC, (2008) in-press.

[11] Berne, R. M and M. N. Levy, ed. *Principles of Physiology*, 3rd ed. St. Louis, MO: Mosby Year Book, Inc., (1993).

[12] Morbiducci, U., L. Scalise, M. De Melis, and M. Grigioni, "Optical Vibrocardiography: A Novel Tool for the Optical Monitoring of Cardiac Activity," *Annals of Biomedical Engineering*, 35(1): 45-58 (2007).

[13] Scalise, L. and U. Morbiducci, "Non-Contact Cardiac Monitoring from Carotid Artery Using Optical Vibrocardiography," *Medical Engineering and Physics*, 30: 490-497 (2008).

[14] Aaronson, Philip I. and Jeremy P. T. Ward, *The Cardiovascular System At A Glance*, edited by MA Malden: Blackwell Publishing Ltd (2004).

[15] Johnston, W. and Y. Mendelson, "Extracting Heart Rate Variability from a Wearable Reflectance Pulse Oximeter." Paper presented at the Proceedings of the IEEE International Conference of the Engineering in Medicine and Biology Society (2004).

[16] Electrophysiology, Task Force of The European Society of Cardiology and The North American Society of Pacing and "Heart Rate Bariability: Standards of Measurement, Physiological Interpretation, and Clinical Use," *European Heart Journal*, 17: 354-381 (1996).

[17] Pan, J. and W. J. Tompkins, "Real Time QRS Detector Algorithm." In *IEEE Transactions on Biomedical Engineering*, 230 (1985).

[18] Greneker, E. F. and J. L. Geisheimer, "The Radar Flashlight Three Years Later: An Update on Developmental Progress." In *Proceedings of the IEEE International Conference on Security Technology*, 257 (2000).

[19] Garbey, M., N. Sun, A. Merla, and I. Pavlidis, "Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery." Paper presented at the IEEE Transactions on Biomedical Engineering,(2007).

[20] "Nasal Cannula", Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki /Nasal_cannula

[21] Rohrbaugh, J. W., E. J. Sirevaag, J. A. Stern, and A. H. Ryan, "The Physiology of Threat: Remote Assessment using Laser Doppler Vibrometry," *Journal of Credibility Assessment and Witness Psychology*, 7(2):
135-145 (2006).

[22] Michahelles, F., R. Wicki, and B. Schiele, "Less Contact: Heart-Rate Detection without Even Touching the User." Paper presented at the Proceedings of the 8th International Symposium on Wearable Computers, October 31-November 3, (2004).

[23] Greneker, E. F., "Radar Sensing of Heartbeat and Respiration at a Distance with Applications on the Technology." Paper presented at the Radar '97, 150-154 (1997).

[24] Geisheimer, J. and E. F. Greneker, "Remote Detection of Deception using Radar Vital Signs Monitor Technology." Paper presented at the IEEE International Conference on Security Technology (2000).

[25] Dowdall, J., I. T. Pavlidis and P. Tsiamyrtzis, "Coalitional Tracking," *Computer Vision and Image Understanding*, **106**, 205-219 (2007).

[26] Fei, J. and I. Pavlidis, "Virtual Thermistor." Paper presented at the Proceedings of the IEEE Engineering in Medicine and Biology Society 29th annual conference, August 22-26 (2007).

[27] Fei, J., Z. Zhu and I. Pavlidis, "Imaging Breathing Rate in the CO2 Absorption Band." Paper presented at the Proceedings of the IEEE Engineering in Medicine and Biology Society 27th annual conference, 700-705 (2005).

[28] Haynes, J. D., and G. Rees, "Decoding Mental States from Brain Activity in Humans," *Nature Reviews Neuroscience*, **7**, 523-534 (2006).

[29] Knight, J., "The Truth About Lying," **Nature**, **428**, 692-694 (2004).

[30] Kozel, F. A., et al., "A Pilot Study of Functional Magnetic Resonance Imaging Brain Correlates of Deception in Healthy Young Men," *Journal Neuropsychiatry and Clinical Neuroscience* **16**, 295-305 (2004).

[31] Langleben, D. D., et al., "Brain Activity During Simulated Deception: An Event-Related Functional Magnetic Resonance Study," *NeuroImage*, **15**, 727-732 (2002).

[32] Levine, J. A., I. Pavlidis and M. Cooper, "The Face of Fear," *The Lancet*, **357**, 1757 (2001).

[33] Lo, Y. L., S. Fook-Chong and E. K. Tan, "Increased Cortical Excitability in Human Deception," *NeuroReport*, **14**, 1021-1024 (2003).

[34] Murthy, R. and I. Pavlidis, "Noncontact Measurement of Breathing Function," *IEEE Engineering In Medicine and Biology*, May/June issue, 57-67 (2006)

[35] Pavlidis, I., et al., "Interacting with Human Physiology," *Computer Vision and Image Understanding*, **108**, 150-170 (2007).

[36] Pavlidis, I., N. L. Eberhardt and J. A. Levine, "Seeing Through the Face of Deception," *Nature*, **415**, 35 (2002a).

[37] Pavlidis, I., N. L. Eberhardt and J. A. Levine, "Seeing Through the face of Deception, Errata," *Nature*, **415**, 602 (2002b).

[38] Pavlidis, I. and J. Levine, "Thermal Analysis for Polygraph Testing," *IEEE Engineering in Medicine and Biology*, November/December issue, 56-64 (2002).

[39] Pollina, D. A., et al., "Facial Skin Surface Temperature Changes during a "Concealed Information" Test," *Annals of Biomedical Engineering*, **34**, 1182-1189 (2006).

[40] Stewart, M., et al., "Non-Invasive Measurement of Stress in Dairy Cows using Infrared Thermography," *Physiology & Behavior*, **92**, 520-525 (2007)

[41] Sun, N., I. Pavlidis, M. Garbey and JFei, "Harvesting the Thermal Car- diac Pulse Signal." In *Medical Image Computing and Computer-Assisted Intervention* edited by M. Nielsen, J. Sporring, and R. Larsen), 569-576 (2006).

[42] Tsiamyrtzis, P., et al., "Imaging Facial Physiology for the Detection of Deceit," *International Journal of Computer Vision*,**71** 197-214 (2006).

[43] Yarbus, A. L ., *Eye Movements and Vision*, Translated by Basil Haigh, New York, NY: Plenum (1967)

[44] Ekman, P. and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Palo Alto, CA: Consulting Psychologists Press (1978).

[45] Ekman, P. et al., "Universal and Cultural Differences in Judgments of Facial Expressions of Emotion," *Journal of Personality and Social Psychology* 53, 712-717, (1987).

[46] Hazlett, G., "Research on Detection of Deception: What We Know vs. What we Think We Know." In *Educing Information: Interogation: Science and Art - Foundations for the Future, Intelligence Science Board, Phase 1 Report,* edited by Russell Swenson, 45-61, Washington, DC: National Defense Intelligence College (2006).

[47] Hadden, D., S. Walter, R. Ratley, and M. Smith, "Investigation and Evaluation of Voice Stress Analysis Technology." Unpublished report to the U.S. Dept. of Justice, (2001).

[48] Hopkins, C. S., R. J. Ratley, D. S. Benincasa, and J. J. Grieco, "Evaluation of Voice Stress Analysis Technology." Paper presented at the Hawaii International Conference on System Sciences (2005).

[49] Damphousse, K. R., L. Pointon, D. Upchurch, and R. K. Moore, "Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting." Report to the U.S. Dept. of Justice (2007).

[50] Cohn, J. F., and P. Ekman, "Measuring Facial Action by Manual Coding, Facial EMG, and Automated Facial Image Analysis." In *The New Handbook of Methods in Nonverbal Behavior Research* edited by R. Rosenthal, J. A. Harrigan, K. R. Scherer, New York: Oxford University Press (2005).

[51] Bartlett, M., et al., "Automatic Recognition of Facial Actions in Spontaneous Expressions," *Journal of Multimedia* 1:22-35 (2006).

[52] Whitehill, J., M. Bartlett, and J. Movellan, "Measuring the Perceived Difficulty of a Lecture using Automatic Facial Expression Recognition," http://mplab.ucsd.edu/ jake/its08.pdf (2008).

[53] Nagl, J. A. and P. J. Schoomaker, *Counterinsurgency Lessons from Malaya and Vietnam: Learning to Eat Soup with a Knife*, Chicago: U. Chicago Press (2005).

[54] Nagl, J. A., D. H. Petreaus, J. F. Amos and S. Sewall, *The U.S. Army/Marine Corps Counterinsurgency Field Manual*, Chicago: U. Chicago Press (2007).

[55] Fisher, R. P., and R. E. Geiselman, *Memory Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*, Springfield, IL.: Charles C. Thomas (1992).

[56] Geiselman, R. E., and R. P. Fisher, "Ten Years of Cognitive Interviewing." In *Intersections in Basic and Applied Memory Research*, edited by D. Payne and F. Conrad, Mahwah, NJ: Lawrence Erlbaum Associates (pp.291-310) (1997).

[57] Fisher, R. P., R. E. Geiselman, and M. Amador, "Field Test of the Cognitive Interview: Enhancing the Recollection of Actual Victims and Witnesses of Crime," *Journal of Applied Psychology*, **74** (5), 722-727 (1989).

[58] Fisher, R. P., R. E. Geiselman, and D. S. Raymond, "Critical Analysis of Police Interviewing Techniques," *Journal of Police Science and Administration* 15, 177-185 (1987).

[59] Geiselman, R. E., R. P. Fisher, D. p. MacKinnon, and H. L. Holland, "Eyewitness Memory Enhancement in the Police Interview: Cognitive Retrieval Mnemonics versus Hypnosis," *Journal of Applied Psychology*, 70, 401-412 (1985).

[60] DeMaio, T. J., and J. M. Rothgeb, "Cognitive Interviewing Techniques: In the Lab and in the Field." In *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, edited by N. Schwarz and S. Sudman, pp. 177-195, San Francisco: Jossey-Bass (1996).

[61] Jabine, T. B., M. L. Straf, J. M. Tanur, and R. Tourangeau, *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, Washington, DC: National Academy Press (1984).

[62] Lessler, J. T., and J. R. Rothgeb, "Integrating Cognitive Research into Household Survey Design." Paper presented at the CASM II Seminar, Hyattsville, Maryland (1999).

[63] Lessler, J. T., and M. G. Sirken, "Laboratory-Based Research on the Cognitive Aspects of Survey Methodology: The goals and methods of the National Center for Health Statistics Study," *Milbank Memorial Fund Quarterly/Health and Society*, 63, 565-581 (1985).

[64] Sirken, M., G., et al. "Cognition and Survey Research," New York: Wiley (1999).

[65] Willis, G., T. DeMaio, and B. Harris-Kojetin, "Is the Bandwagon Headed to the Methodological Promised Land? Evaluation of the Validity of Cognitive Interviewing Techniques." In *Cognition and Survey Research*, edited by D. Herrmann, M. sirken, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau New York: Wiley (1999).

[66] Willis, G.B., and S. Schechter, "Evaluation of Cognitive Interviewing Techniques: Do the Results Generalize to the Field?" *Bulletin de Methodologie Sociologique* 55, 40-66 (1997).

[67] Vrij, A., R. Fisher, S. Mann, and S. Leal "Lie Detection: Pitfalls and Opportunities." In *Interrogations and Confessions: Current Research, Practice and Policy* edited by D. Lassiter & C. Meissner, Washington, DC: American Psychological Association, in press.

[68] Fisher, R., "Cognitive Interviews for Deception," personal correspondence, July 16, 2008.

[69] Feynman, R., "Cargo Cult Science," *Engineering and Science*, Vol. 37:7, 10-13 (1974).

[70] Bausell, R. B., *Snake Oil Science: The Truth About Complementary and Alternative Medicine*, New York: Oxford University Press, (2007).

[71] Theile, G., "Wave Field Synthesis a Promising Spatial Audio Rendering Concept." Paper presented at the International Conference on Digital Audio Effects, Naples, Italy October 5-8 (2004).

[72] Thalmann, D., "Using Virtual Reality Techniques in the Animation Process." In *Virtual Reality Systems*, edited by M. Gigante, R. Earnshaw, H. Jones, Academic Press, pp. 143-159 (1993).

[73] Livingston, M. A. and A. State, "Magnetic Tracker Calibration for Improved Augmented Reality Registration," *PRESENCE: Teleoperators and Virtual Environments*, 6: 532-546 (1997).

[74] Ghazisaedy, M., et al., "Ultrasonic Calibration of a Magnetic Tracker in a Virtual Reality Space." Paper presented at the Virtual Reality Annual International Symposium (VRAIS '95) (1995).

DISTRIBUTION LIST

Assistant Secretary of the Navy
(Research, Development & Acquisition)
1000 Navy Pentagon
Washington, DC 20350-1000

Assistant Deputy Administrator for
Military Application
NA-12
National Nuclear Security Administration
U.S. Department of Energy
1000 Independence Avenue, SW
Washington, DC 20585

DARPA Library
3701 North Fairfax Drive
Arlington, VA 22203-1714

Deputy Under Secretary of
Defense Science & Technology
3040 Defense Pentagon
Washington, DC 20301-3040

Defense Technical Information Center (DTIC)
8725 John J. Kingman Road
ATTN: DTIC-OA (Mr. Jack Rike)
Suite 0944
Fort Belvoir, VA 22060-6218

Director, DTRA
Research Development Office
8725 John Jay Kingman Road
Room 3380, Mail Stop 6201
Fort Belvoir, VA 22060

Headquarters Air Force XON
4A870 1480 Air Force Pentagon
Washington, DC 20330-1480

IC JASON Program [2]
Chief Technical Officer/OCS
2P0104 NHB
Central Intelligence Agency
Washington, DC 20505-0001

Records Resource
The MITRE Corporation
Mail Stop D460
202 Burlington Road, Rte 62
Bedford, MA 01730-1420

Reports Collection
Los Alamos National Laboratory
Mail Station 5000, MS A150
PO Box 1663
Los Alamos, NM 87545

Superintendent
Code 1424
Attn: Documents Librarian
Naval Postgraduate School
Monterey, CA 93943

U S Army Space & Missile Defense Command
Attn: SMDC-ZD (Dr. Swinson)
PO Box 1500
Huntsville, AL 35807-38017

Administrator
U.S. Dept of Energy
National Nuclear Security Administration
1000 Independence Avenue, SW
NA-10 FORS Bldg
Washington, DC 20585

Principal Deputy Director
Office of Science, SC-2/Forrestal Building
U.S. Department of Energy
1000 Independence Avenue, SW
Washington, DC 20585

Director
Institute for Defense Analyses
Technical Information Services
Room 8701
4850 Mark Center Drive
Alexandria, VA 22311-1882

JASON Library [5]
The MITRE Corporation
3550 General Atomics Court
Building 29
San Diego, CA 92121-1122

Dr. Lawrence K. Gershwin
NIC/NIO/S&T
2E42, OHB
Washington, DC 20505

Dr. Alfred Grasso
President & CEO
The MITRE Corporation
Mail Stop N640
7515 Colshire Drive
McLean, VA 22102-7508

Dr. Barry Hannah
Reentry Systems Branch Head, Navy Strategic Systems Programs
Strategic Systems Programs (Attn: SP28)
2521 Clark Street, Suite 1000
Arlington, VA 22202-3930

Dr. Robert G. Henderson
The MITRE Corporation
Mailstop MDA/ Room 5H305
7515 Colshire Drive
McLean, VA 22102-7508

Dr. Bobby R. Junker
Office of Naval Research
Code 31
800 North Quincy Street
Arlington, VA 22217-5660

Mr. Kevin "Spanky" Kirsch
Director, Special Programs
US Department of Homeland Security
Science and Technology Directorate
Washington, DC 20528

Dr. Daniel J. McMorrow
Director, JASON Program Office
MITRE Corporation Mailstop
T130
7515 Colshire Drive
McLean, VA 22102-7508

Dr. A. Fenner Milton
Director, CERDEC NVESD
10221 Burbeck Road
Ft. Belvoir, VA 22060-5806

Dr. Julian C. Nall
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

Deputy Chief Scientist
U.S. Army Space & Missile Defense Command
PO Box 15280
Arlington, VA 22215-0280

Dr. John A. Parmentola
Director, Army Research and Labs
Presidential Tower, Suite 98C
2511 Jefferson Davis Highway
Arlington, VA 22202

Dr. John R. Phillips
Chief Scientist, DST/CS
2P0104 NHB
Central Intelligence Agency
Washington, DC 20505-0001

Dr. William S. Rees, Jr.
OSD/DDR&E
Deputy Under Secretary of Defense for Laboratories and Basic Sciences
3040 Defense Pentagon
Room 3C 913A

Washington, DC 20301-3040

Dr. Alan R. Shaffer
Principal Deputy Director
DDR&E
3040 Defense Pentagon
Room 3B 854
Washington, DC 20301-3040

Mr. Anthony J. Tether
DIRO/DARPA
3701 N. Fairfax Drive
Arlington, VA 22203-1714