Behavioral Science and Security

Statement of Philip E. Rubin, Ph.D.

before the
Subcommittee on Investigations and Oversight
Committee on Science, Space, and Technology
U.S. House of Representatives

April 6, 2011

Chairman Broun, Ranking Member Edwards, and members of the Subcommittee on Investigations and Oversight of the Committee on Science, Space, and Technology, thank you for the opportunity to speak to you today. My name is Philip Rubin, a resident of Fairfield, Connecticut. I am here as a private citizen. However, I currently serve or have served in a number of roles, both inside and outside of government, that might be relevant to today's hearing. In addition to the separate biography and resume that I have provided, I will mention some key positions and/or responsibilities. I am the Chief Executive Officer and a senior scientist at Haskins Laboratories in New Haven, Connecticut, a private, non-profit research institute affiliated with Yale University and the University of Connecticut that has a primary focus on the science of the spoken and written word, including speech, language, and reading, and their biological basis. I am also an adjunct professor in the Department of Surgery, Otolaryngology at the Yale University School of Medicine. My research spans a number of disciplines, combining computational, engineering, linguistic, physiological, and psychological approaches to study embodied cognition, most particularly the biological bases of speech and language.

Since 2006 I have served as the Chair of the National Academies Board on Behavioral, Cognitive, and Sensory Sciences. I was also the Chair of the National Research Council (NRC) Committee on Field Evaluation of Behavioral and Cognitive Sciences-Based Methods and Tools for Intelligence and Counter-Intelligence, and a member of the NRC Committee on Developing Metrics for Department of Homeland Security Science and Technology Research. I am a member-at-large of the Executive Committee of the Federation of Associations in Behavioral & Brain Sciences. The American Institutes for Research (AIR), at the request of the Department of Homeland Security Science & Technology, is conducting a study to assess the validity of the

Transportation Security Administration's (TSA) Screening of Passengers by Observation Techniques (SPOT) program's primary instrument, the SPOT Referral Report, to identify "high risk travelers." I am a member of the Technical Advisory Committee (TAC) that was formed to provide critical input related to analyses and methodologies in this project. The final report is expected shortly. The SPOT review is an ongoing activity and I have let this committee's staff know that I have signed a nondisclosure agreement about aspects of the program. Since Feb. 2011 I have also been a member of the federal interagency High-Value Detainee Interrogation Group (HIG) Research Committee. From 2000 through 2003 I served as the Director of the Division of Behavioral and Cognitive Sciences at the National Science Foundation (NSF). During that period I served as the co-chair of the interagency NSTC Committee on Science Human Subjects Research Subcommittee under the auspices of the Executive Office of the President, Office of Science and Technology Policy (OSTP) during both the Clinton and Bush administrations. I was also a member of the NSTC Interagency Working Group on Social, Behavioral and Economic Sciences Task Force on Anti-Terrorism Research and Development during the Bush administration.

I was invited here today to describe the current state of research and science in the behavioral and cognitive sciences related to laboratory studies and field evaluation of various tools, techniques, and technologies used in security and the detection of deception. My testimony will summarize some activities in these areas, particularly those with which I have personal experience, that might be of use to this subcommittee.

Before describing some recent reports of significance, let me begin by noting some activities of particular relevance to behavioral science and security. The significance of the behavioral and cognitive sciences to matters of security was

highlighted within the intelligence community in a number of articles written from 1978 to 1986 by Richards J. Heuer, Jr., an analyst with the Central Intelligence Agency. These were later collected in a book, *Psychology of Intelligence Analysis* (Heuer, 1999), that surveyed cognitive psychology literature and suggested ways to apply these research findings to improve performance in various tasks.

On Feb. 10, 2005, The National Science and Technology Council (NSTC) released the report "Combating Terrorism: Research Priorities in the Social, Behavioral and Economic Sciences." Produced by the Subcommittee on Social, Behavioral and Economic Sciences, this was the first NSTC report on the role of the social and behavioral sciences (which include psychology, sociology, anthropology, geography, linguistics, statistics, and statistical and data mining) in helping the American public and its leaders to understand the causes of terrorism and how to counter terrorism. As a member of the NSTC Interagency Working Group on Social, Behavioral and Economic Sciences Task Force on Anti-Terrorism Research and Development, I was one of the individuals who helped to draft the initial versions of this report. The focus of the report was on how these sciences can help us to predict, prevent, prepare for and recover from a terrorist attack or ongoing terrorists' threats. A revised, printed form of the report was released in 2009. Speaking of this report, John H. Marburger III, then science advisor to the President and director of the Office of Science and Technology Policy, said, "Our ability to maintain our American way of life depends on our understanding of human behavior, which is the domain of the social, behavioral and economic sciences. The report describes the powerful tools and strategies these sciences offer as we respond to the threats and actions of terrorists." The report goes on to say, in part, that:

> "Terrorism has enormous impacts beyond the immediate destruction, injury, loss of life, and consequent fear and panic. These impacts span the personal,

> organizational and societal levels and can have profound psychological, economic and social consequences. They apply not just to terrorist activity, but to other crises of national and/or regional import, such as natural disasters, industrial accidents, and other extreme events. Research in the social, behavioral and educational sciences has also provided the knowledge, tools, techniques, and trained scientists that are needed if we are to be prepared to understand, prevent, mitigate, and intervene where required in events related to such national crises. Lessons learned from previous research and development efforts are diverse and numerous. For example, research on the mental health consequences of disasters, including terrorist acts such as the Oklahoma City bombing, has produced a better understanding of the course of disruptive and disabling symptoms of distress, who is at risk of developing a serious mental illness, and helpful interventions to reduce trauma-related distress including depression and anxiety disorders. Basic economic research on how markets work was used by government economic advisors to devise policies that would provide the right incentives and not interfere with transitions in industries most affected by the changed security situation after 9/11."

Other important work related to the behavioral sciences and security included work by the Intelligence Science Board on the art and science of interrogation, described in the volume *Educing Information* (2006). Rapid developments in cognitive neuroimaging technologies (PET, fMRI, MEG, NIRS, EEG, etc.) and their possibility use in the detection of deception, attitude, and affect, have led to the beginnings of a cottage industry in what some have called "brain reading" or "brain fingerprinting." In his 2006 book, *Mind Wars: Brain Research and National Defense*, Jonathan Moreno, discusses current concerns related to such developments.

"It's especially hard to assess the plausibility that something such as mind reading or mind control is feasible through the kinds of devices I've described … Many of the technologies do seem hyped; just because national security agencies are spending money on them doesn't mean they are a sure thing … With brain theory as inconclusive as it is, there are bound to be conflicting claims among neuroscientists about what's technically possible and what isn't. Since neuroscience hasn't come close to finding the boundaries of its possibilities yet, that uncertainty is likely to persist for a long time." (112-113)

Things change rapidly in science and technology, however as recently as this month one of our leading cognitive neuroscientists, Michael Gazzaniga, while enthusiastic about the potential of work in the area, struck a note of caution in an article in *Scientific American* (April 2011) called "Neuroscience in the Courtroom." Speaking from a legal perspective related to the admissibility of juvenile brain scans as evidence, he said, "In spite of the many insights pouring forth from neuroscience, recent findings from research into the juvenile mind highlight the need to be cautious when incorporating such science into the law." … "Exciting as the advances that neuroscience is making everyday are, all of us should look with caution at how they may gradually become incorporated into our culture. The legal relevance of neuroscientific discoveries is only part of the picture."

The National Academies, comprised of the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and their operating arm, the National Research Council, provide independent, objective advice and supporting information on issues that affect all of our citizens' lives. This takes a number of forms, including published documents such as consensus reports, workshop summaries, and paper collections. A number of these are of particular relevance to today's hearing, and I will list or summarize the most important ones. Most of these were produced under the supervision of the Division of Behavioral and Social Sciences and Education (DBASSE) of the NRC and the Board on Behavioral, Cognitive, and Sensory Sciences (BBCSS) that I chair. Since its founding in 1997, BBCSS has developed and managed many major studies conducted by expert panels, involving hundreds of volunteers including scientists, policymakers, government employees, and public citizens. The goal has been to create a sustainable infrastructure for ongoing review of fundamental and translational research,

to inform policy on issues of national priority, and to facilitate interactions among scholars and policymakers. Meetings and activities of BBCSS have been sponsored, in part, by: the National Science Foundation, Directorate for Social, Behavioral and Economic Sciences; the National Institutes of Health, including the National Institute on Aging, Division of Behavioral and Social Research, the National Cancer Institute; and the Office of Behavioral and Social Science Research (OBSSR); the American Psychological Association; the Office of the Director of National Intelligence (ODNI); the Defense Intelligence Agency (DIA); and the U. S. Secret Service. For today's purposes, the most relevant documents include:

The Polygraph and Lie Detection. (2003)

Human Behavior in Military Contexts. (2008)

Behavioral Modeling and Simulation: From Individuals to Societies. (2008)

Emerging Cognitive Neuroscience and Related Technologies. (2008)

Protecting Individual Privacy in the Struggle Against Terrorists. (2008)

Field Evaluation in the Intelligence and Counterintelligence Context. (2010)

Intelligence Analysis: Behavioral and Social Scientific Foundations. (2011)

Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences. (2011)

Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures. (2011)

Time and space prevent a detailed description of these important documents. Instead I will focus on the *Field Evaluation* workshop summary and *Threatening Communications* paper collection.

Field Evaluation

On September 22-23, 2009, the Board on Behavioral, Cognitive, and Sensory Sciences of the NRC held a workshop on the field evaluation of behavioral and cognitive sciences–based methods and tools for use in the areas of intelligence and counterintelligence. The workshop was organized by the Planning Committee on Field Evaluation of Behavioral and Cognitive Sciences-Based Methods and Tools for Intelligence and Counterintelligence that I chaired. Its purpose was to discuss the best ways to apply methods and tools from the behavioral sciences to work in intelligence operations. The workshop focused on the issue of *field evaluation*—the testing of these methods and tools in the context in which they will be used in order to determine if they are effective in real-world settings. The workshop was sponsored by the DIA and the ODNI and had considerable support from Susan Brandon, then chief for research, Behavioral Science Program DEO– Defense CI and HUMINT Center DIA, and Steven Rieber, then research director, Office of Analytic Integrity and Standards, ODNI.

In 2010, the NRC published a workshop summary called *Field Evaluation in the Intelligence and Counterintelligence Context*. This short document summarized the meeting and highlighted key issues. Following [the single-spaced sections] are quoted extracts of the *Field Evaluation* Workshop Summary, with minor edits for continuity [attribution quotes are omitted], that detail some of these issues and illustrate weaknesses in our current approaches, while also considering future opportunities.

In one of the workshop presentations, David Mandel, a senior defense scientist at Defence Research and Development Canada (DRDC), discussed the ways in which the behavioral sciences can benefit intelligence analysis and why it is important for the intelligence community to build a partnership with the behavioral sciences community. The intelligence community has long relied on science and technology for insights and

techniques, Mandel noted, so one might wonder why it is necessary to talk about the importance of strengthening the relationship between the intelligence community and the broad community of behavioral scientists. One important reason, he said, is that there are a number of factors that tend to weaken the relationship between the two communities and make analysts less likely to take advantage of what the behavioral sciences can offer. First, Mandel said, there is a natural inclination among most people— including those in the intelligence community—to react poorly to "scholarly verdicts that deal with issues such as the quality of their judgment and decision making, their susceptibility to irrational biases, their use of suboptimal heuristics, and overreliance on non-diagnostic information." Like most people, experts have the sense that they are competent. Psychological research shows that most people believe themselves to be better than average at what they do. Thus, Mandel said, experts are prone to challenge conclusions offered by behavioral scientists with their own knowledge gained from personal experience and, furthermore, to believe that such a challenge is completely legitimate. This is a fundamental problem that behavioral scientists face in making contributions to any practitioner community, Mandel said, "Their research is very easily disregarded on the basis of intuition and common sense. A second reason that analysts tend to disregard lessons from behavioral science is that it is seen as being "soft" science. Thus its knowledge is considered to be less objective or trustworthy than knowledge generated by the "hard" sciences and technology, such as satellite imaging or electronic eavesdropping. Although that attitude is common in the intelligence community, Mandel cautioned, it is misguided and underestimates both the value and the analytical power of behavioral science. "When someone uses the term 'soft science,' I correct them. I say 'probabilistic science' and [note that] we deal with some very difficult problems." Third, Mandel said, the relationship between the intelligence community and the behavioral science community is still relatively new, so analysts do not necessarily understand what behavioral science has to offer. Thus, he noted, forums like this workshop are important for exploring ways in which the partnership between the two communities can be developed.

It is telling, Mandel noted, that no one else has come along since Heuer to continue his work of translating cognitive psychology and other areas of behavioral science into tools for analysis. In cognitive psychology alone there is at least a quarter century of new research since Heuer published *Psychology of Intelligence Analysis* that is waiting to be exploited by the intelligence community. Another way in which establishing a connection with the research community can help the intelligence community is with validation, Mandel said. Once knowledge and insights from behavioral science are used to develop new tools for the intelligence community, it is still necessary to validate them. Simply basing recommendations on scientific research is not the same thing as showing scientifically that those recommendations are effective or testing to see if they could be substantially improved. Even Heuer was unable to do much to validate his recommendations, Mandel noted, and, more generally, this is not something that the intelligence community is particularly well equipped to do. It is, however, exactly what research scientists are trained to do. Science offers a method for testing which ideas lead to good results and which do not. Thus, partnering with the behavioral science community can help the intelligence community zero in on the techniques that work best and avoid those that work poorly or not at all.

In theory, Mandel said, it would be possible for the intelligence community to build its own applied behavioral research capability, but that would draw significant resources away from other operational areas and add an entirely new focus and purpose to the intelligence community's existing tasks. Furthermore, if the intelligence community were to hire behavioral scientists, it would find itself in competition with both academia, with its unparalleled freedoms, and industry, with its lucrative salaries. It makes more sense, Mandel suggested, for the intelligence community to develop partnerships with universities and other institutions that already have the expertise and capability to perform behavioral science research. A final advantage of partnering with the existing behavioral science community, Mandel said, is the "multiplier effect." By working with scientists in academia, for example, the intelligence community is not only drawing on the knowledge of those subject-matter experts but on all of their contacts. "As a researcher in a research and development organization and government," Mandel said, "I am very keen on partnering with academics because I understand that they have the ability to reach back into other areas of academia and connect me with other experts who could be of use." There is a tremendous amount of such leverage that can be achieved by building relationships rather than trying to do everything in-house.

In what ways might particular tools and techniques from the behavioral sciences assist the intelligence and counterintelligence community? A variety of devices and approaches derived from the behavioral sciences have been suggested for use or have already been used by the intelligence community. Several of these were described, with a particular emphasis on how the techniques have been evaluated in the field. As Robert Fein put it, "Our spirit here is to move forward, to figure out what kinds of new ideas, approaches, old ideas might be useful to defense and intelligence communities as they seek to fulfill what are often very difficult and sometimes awesome responsibilities." To that end the speakers provided case studies of various technologies with potential application to the intelligence field. One common thread among all of these disparate techniques, a point made throughout the workshop, is that none of them has been subjected to a careful field evaluation.

Deception Detection

People in the military, in law enforcement, and in the intelligence community regularly deal with people who deceive them. These people may be working for or sympathize with an adversary, they may have done something they are trying to hide, or they may simply have their own personal reasons for not telling the truth. But no matter the reasons, an important task for anyone gathering information in these arenas is to be able to detect deception. In Iraq or Afghanistan, for example, soldiers on the front line often must decide whether a particular local person is telling the truth about a cache of explosives or an impending attack. And since research has shown that most individuals detect deception at a rate that is little better than random chance, it would be useful to have a way to improve the odds. Because of this need, a number of devices and methods have been developed that purport to detect deception. Two in particular were described at the workshop: *voice stress technologies* and the *Preliminary Credibility Assessment Screening System (PCASS)*.

Voice Stress Technologies

Of the various devices that have been developed to help detect lies and deception, a great many fall in the category of voice stress technologies. I offered a brief overview of these technologies and of how well they have performed on objective tests. The basic idea behind all of these technologies is that a person who answers a question deceptively will feel a heightened degree of stress, and that stress will cause a change in voice characteristics that can be detected by a careful analysis of the voice. The change in the voice may not be audible to the human ear, but the claim is that it can be ascertained accurately and reliably by using signal-processing techniques. More specifically, many of the voice stress technologies are based on the assumption that microtremors—vibrations of such a low frequency that they cannot be detected by the human ear—are normally present in human speech but that when a person is stressed, the microtremors are suppressed. Thus by monitoring the microtremors and noting when they disappear, it should be possible to determine when a person is speaking under stress—and presumably lying or otherwise trying to deceive.

Over the years, these technologies have been tested by various researchers in various ways. A review of these studies that was carried out by Sujeeta Bhatt and Susan Brandon of the Defense Intelligence Agency (Bhatt and Brandon, 2009). After examining two dozen studies conducted over 30 years, the researchers concluded that the various voice stress technologies were performing, in general, at a level no better than chance—a person flipping a coin would be equally good at detecting deception. In short, there was no evidence for the validity or the reliability of voice stress analysis for the detection of deception in individuals. Furthermore not only is there no evidence that voice stress technologies are effective in detecting stress, but also the hypothesis underlying their use has been shown to be false. If indeed there are microtremors in the voice, then they must result from tremors in some part of the vocal tract—the larynx, perhaps, or the supralaryngeal vocal tract, which is everything above the larynx, including the oral and nasal cavities. Using a technique called electromyography to measure the electrical signals of muscle activities, physiologists have found that there are indeed microtremors of the correct frequency—about 8 to 12 hertz—in some muscles, including those of the arm. So it would seem reasonable to think that there might also be such microtremors in the vocal tract, which would produce microtremors in the voice. However, research has found no such microtremors, either in the muscles of the vocal tract or in the voice itself. So the basic idea underlying voice stress technologies—that stress causes the normal microtremors in the voice to be suppressed—is not supported by the evidence.

The claim is not that voice stress technologies do not work, only that there has been extensive testing with very little evidence that such technologies do work. It is possible that some of the technologies do work under certain conditions and in certain circumstances, but if that is so, more careful testing will be needed to determine what those conditions and circumstances are. And only when such testing has been carried out and the appropriate conditions and circumstances identified will it make sense to carry out field evaluations of such technologies. At this point, voice stress technologies are not ready for field evaluation. For the most part the intelligence community has now stayed

away from voice stress technologies mainly because of the absence of any evidence supporting their accuracy. But the law enforcement community has taken a difference approach. Despite the lack of evidence that the various voice stress technologies work, and despite the absence of any field evaluations of them, the technologies have been put to work by a number of law enforcement agencies around the country and around the world. It is not difficult to understand the reasons. The devices are inexpensive. They are small and do not require that sensors be attached to the person being questioned; indeed, they can even be used in recorded sessions. And they require much less training to operate than a polygraph. Many people in law enforcement believe that the voice stress technologies do work; even among those who are convinced that the results of the technologies are unreliable, many still believe that the devices can be useful in interrogations. They contend that simply questioning a person with such a device present can, if the person believes that it can tell the difference between the truth and a lie, induce that person to tell the truth.

Preliminary Credibility Assessment Screening System

With the reliability of voice stress technologies called into question, the intelligence community needed another way to screen for deception. Donald Krapohl, special assistant to the director of the Defense Academy for Credibility Assessment (DACA), described how, several years ago, the Pentagon asked DACA for a summary of the research on voice stress technologies. DACA, which is part of the Defense Intelligence Agency in the Department of Defense, provided a review of what was known about voice stress analysis, and, as Krapohl put it, "it was rather scary to them, and they decided to pull those technologies back."

The need for deception detection remained, however, and DACA's headquarters organization, the Counterintelligence Field Activity (CIFA) (CIFA was shut down in 2008 and its responsibilities were taken over by a new agency, the Defense Counterintelligence and Human Intelligence Center), was given the job of finding a new technology that would do the same job that voice stress technologies were supposed to perform, but with significantly more accuracy. There were a number of requirements in order for a device to be effective in the field: it had to have low training requirements, as it would be used by soldiers on the front line rather than interrogation specialists; ideally it would require no more than a week of training. It needed to be highly portable and easy to use for the average soldier. It needed to be rugged, as inevitably it would be dropped, get wet, and get dirty.

And it had to be a deception test, not a recognition test. That is, instead of recognizing when someone knows something that they are trying to hide—the so-called guilty knowledge test—it should be able to detect when someone was giving a deceptive answer to a direct question. There is a great deal of research concerning the guilty knowledge test, Krapohl explained, but the test is not particularly useful in the field because the interviewers must know something about the "ground truth." Deception tests, by contrast, are not as well understood by the scientific community, but they are far more useful in the field, where interviewers may not know the ground truth.

The final requirement for the device was that it needed to be relatively accurate as an initial screening tool. It was never intended to provide a final answer of whether someone was telling the truth. Its purpose instead was to provide a sort of triage: when

soldiers in the field question someone who claims to have some information, they need to weed out those who are lying. The ones who are not weeded out at this initial stage would be questioned further and in more detail. There are polygraph examiners who can perform extensive examinations, Krapohl explained, but their numbers are limited. "So if you could use a screening tool up front to decide who gets the interview, who gets the interrogation, who gets the polygraph examination, the commanders thought that would be very useful," he said. "It was not designed to be a standalone tool. It was designed only as an initial assessment."

One of the key facts about PCASS is that it was designed specifically to detect deception, which made it possible, Krapohl said, to create an algorithm that considers all of the response data and provides a straightforward answer to the question of whether a person is being deceptive: yes, no, or maybe. It does not provide nearly as much information as a polygraph can, but that is not its purpose. The main use for PCASS is on the front lines where soldiers need help in determining who seems trustworthy and who seems to have something to hide. But the technique is not assumed to give a definite answer, only a conditional one. Because PCASS is used on the front lines, it has never been field tested. Still, it has proved its value in various ways, he said. In a recent operation in Iraq, for example, it allowed U.S. forces to identify a number of individuals who were working for foreign intelligence services and others who were working for violent extremist organizations.

Still, Krapohl said, there is more work to be done. The group at DACA thinks, for example, that by taking advantage of some of the state-of-the- art technologies for deception detection, it should be possible to develop more accurate versions of PCASS. In particular, by using the so-called directed lie approach—in which those being questioned are instructed to provide false answers to certain comparison questions—it should be possible to get greater standardization and less intrusiveness, he said. Still, the issue of field evaluation remains, Krapohl said. Although the technique has been tested in the laboratory, there are no data on its performance in the field. "Doing validation studies of the credibility assessment technology in a war zone has a number of problems that we have not been able to figure out," he said. Nonetheless, DACA researchers would like to come up with ideas for how PCASS and other credibility assessment technologies might be evaluated in the field.

In later discussions at the workshop, it became clear that a number of participants had serious doubts about the effectiveness of PCASS in the field, despite the fact that it is in widespread use and popular among at least some of the troops in the field. "Everybody in this room knows that there are real limitations to it," Fein said. "I think we can do better than put something out there that has such limitations." And Brandon commented that "if we were doing really good field validation with the PCASS" then it might well become obvious that other, less expensive methods could do at least as good a job as PCASS at detecting deception. There are a number of important questions concerning the validity and reliability of PCASS that can be addressed only by field evaluation, and until such validation is done, the troops in the field are relying on what is essentially an unproved technology.

Obstacles To Field Evaluation

A number of the workshop presenters and participants spoke about various obstacles to field evaluation inside the intelligence community— obstacles they believe must be overcome if field evaluation of techniques and devices derived from the behavioral sciences is to become more common and accepted.

Lack of Appreciation of the Value of Field Evaluations

Perhaps the most basic obstacle is simply a lack of appreciation among many of those in the intelligence community for the value of objective field evaluations and how inaccurate informal "lessons learned" approaches to field evaluation can be. Paul Lehner of the MITRE Corporation made this point, for instance, when he noted that after the 9/11 attacks on the World Trade Center there was a great sense of urgency to develop new and better ways to gather and analyze intelligence information—but there was no corresponding urgency to evaluate the various approaches to determine what really works and what doesn't.

David Mandel commented that this is simply not a way of thinking that the intelligence community is familiar with. People in the intelligence and defense communities are accustomed to investing in devices, like a voice stress analyzer, or other techniques, but the idea of field evaluation as a deliverable is foreign to most of them. Mandel described conversations he had with a military research board in which he explained the idea of doing research on methods in order to determine their effectiveness. "The ideas had never been presented to the board," he said. "They use [various techniques], but they had never heard of such a thing as research on the effectiveness of [them]." The money was there, however, and once the leaders of the organization understood the value of the sort of research that Mandel does, he was given ample funding to pursue his studies.

One of the audience members, Hal Arkes of Ohio State University, made a similar point when he said that the lack of a scientific background among many of the staff of executive agencies is a serious problem. "If we have recommendations that we think are scientifically valid or if there are tests done that show method A is better than method B, a big communication need is still at hand," he said. "We have to convince the people who make the decisions that the recommendations that we make are scientific and therefore are based on things that are better than their intuition, or better than the anecdote that they heard last Thursday evening over a cocktail."

A Sense of Urgency to Use Applications and Institutional Biases

A number of people throughout the meeting spoke about the pressures to use new devices and techniques once they become available because lives are at stake. For example, Anthony Veney, chief of counterintelligence investigation and functional services at U.S. Central Command, spoke passionately about the people on the front lines in Iraq and Afghanistan who need help now to prevent the violence and killings that are going on. But, as other speakers noted, this sense of urgency can lead to pressure to use available tools before they are evaluated—and even to ignoring the results of evaluations if they disagree with the users' conviction that the tools are useful.

Robert Fein described a relevant experience with polygraphs. The NRC had completed its study on polygraphs, which basically concluded that the machines have very limited usefulness for personnel security evaluations, and the findings were being

presented in a briefing (National Research Council, 2003). It was obvious, Fein said, that a number of the audience members were becoming increasingly upset. "Finally, one gentleman raised his hand in some degree of agitation, got up and said, 'Listen, the research suggests that psychological tests don't work, the research suggests that background investigations don't work, the research suggests interviews don't work. If you take the polygraph away, we've got nothing.'" A year and a half later, Fein said, he attended a meeting of persons and organizations concerned with credibility assessment, at which one security agency after another described how they were still using polygraph testing for personnel security evaluations as often as ever. It seemed likely, Fein concluded, that the meticulously performed study by the NRC had had essentially no effect on how often polygraphs were used for personnel security.

The reason, suggested Susan Brandon, is that people want to have some method or device that they can use, and they are not likely to be willing to give up a tool that they perceive as useful and that is already in hand if there is nothing to replace it. This was probably the case, she said, when the U.S. Department of Defense decided to stop using voice stress analysis–based technologies because the data showed that they were ineffective. The user community had thought they were useful, and when they were taken away, a vacuum was left. The users of these technologies then looked around for replacement tools. The problem, Brandon said, is that the things that get sucked into this vacuum may be worse than what they were replacing. So those doing field evaluations must think carefully about what options they can offer the user community to replace a tool that is found ineffective.

I offered a similar thought. The people in the field often do not want to wait for further research and evaluation once a technology is available and there are those out there that will exploit some of these gray areas and faults and will try to sell snake oil to us. The question is, How to push back? How to prevent the use of technology that has not been validated, given the sense of urgency in the intelligence field? And how does one get people in the field to understand the importance of validation in the first place? These are major concerns. Some of the most intractable obstacles to performing field evaluations of intelligence methods are institutional biases. Because these can arise even when everyone is trying to do the right thing, such biases can be particularly difficult to overcome.

Threatening Communications

In March 2011, the NRC released a small collection of papers on the subject of threatening communications and behavior. In my introduction (along with Barbara A. Wanchisen) to the volume, we say:

"Today's world of rapid social, technological, and behavioral change provides new opportunities for communications with few limitations of time or space. The ease by which communications can be made with- out personal proximity has dramatically affected the volume, types, and topics of communications between individuals and groups. Through these communications, people leave behind an ever-growing collection of traces of their daily activities, including digital footprints provided by text, voice, and

other modes of communication. Many personal communications now take place in public forums, and social groups form between individuals who previously might have acted in isolation. Ideas are shared and behaviors encouraged, including threatening or violent ideas and behaviors. Meanwhile, new techniques for aggregating and evaluating diverse and multimodal information sources are available to security services that must reliably identify communications indicating a high likelihood of future violence."

The papers reviewed the behavioral and social sciences research on the likelihood that someone who engages in abnormal and/or threatening communications would actually then try to do harm. They focused on "how scientific knowledge can inform and advance future research on threat assessments, in part by considering the approaches and techniques used to analyze communications and behavior in the dynamic context of today's world. Authors were asked to present and assess scientific research on the correlation between communication-relevant factors and the likelihood that an individual who poses a threat will act on it. The authors were encouraged to consider not only communications containing direct threats, but also odd and inappropriate communications that could display evidence of fixation, obsession, grandiosity, entitled reciprocity, and mental illness."

"The papers in this collection were written within the context of protecting high-profile public figures from potential attack or harm. The research, however, is broadly applicable to U.S. national security including potential applications for analysis of communications from leaders of hostile nations and public threats from terrorist groups. This work high- lights the complex psychology of threatening communications and behavior, and it offers knowledge and perspectives from multiple domains that can contribute to a deeper understanding of the value of communications in predicting and preventing violent behaviors."

This volume focused on communication, forensic psychology, and the analysis of language-based datasets (corpora) to help identify and understand threatening

communications and responses to them through text analysis. It serves as an example of the kind of synthesis of current knowledge that is useful for generating ideas for potential new research directions. (Chung & Pennebaker, 2011; Meloy, 2011; O'Hair, et al, 2011).

TSA's SPOT program

The United States Government Accountability Office's (GAO) May 2010 report, "Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges," questioned whether there was a scientifically valid basis for using behavior and appearance indicators as a means for reliably identifying passengers who may pose a risk to the U.S. aviation system. The report said that, "According to TSA, SPOT was deployed before a scientific validation of the program was completed in response to the need to address potential threats, but was based upon scientific research available at the time regarding human behaviors. TSA officials also stated that no other large-scale U.S. or international screening program incorporating behavior- and appearance-based indicators has ever been rigorously scientifically validated." The GAO report also mentioned a separate report by the JASON group ("The Quest for Truth: Deception and Intent Deception") that had significant concerns about the SPOT program.

The GAO pointed out that a 2008 NRC report indicated that information-based programs, such as behavior detection programs, should first determine if a scientific foundation exists and use scientifically valid criteria to evaluate its effectiveness before going forward. "The report added that programs should have a sound experimental basis and that the documentation on the program's effectiveness should be reviewed by an

independent entity capable of evaluating the supporting scientific evidence. Thus, and as recommended in GAO's May 2010 report, an independent panel of experts could help DHS develop a comprehensive methodology to determine if the SPOT program is based on valid scientific principles that can be effectively applied in an airport environment for counterterrorism purposes. Specifically, GAO's May 2010 report recommended that the Secretary of Homeland Security convene an independent panel of experts to review the methodology of a validation study on the SPOT program being conducted by DHS's Science and Technology Directorate to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program. GAO recommended that this assessment include appropriate input from other federal agencies with expertise in behavior detection and relevant subject matter experts. DHS concurred and stated that its current validation study includes an independent review of the program that will include input from other federal agencies and relevant experts." According to DHS, this independent review is expected to be completed soon.

As indicated above, I am a member of the Technical Advisory Committee (TAC) for SPOT. As the GAO report indicates, TAC's role is extremely limited, focusing in the main on determining whether or not the research program successfully accomplished the goal of evaluating whether SPOT can identify "high-risk travelers" (i.e., individuals who are knowingly and intentionally attempting to defeat the airport security process). TAC has not been asked to evaluate the overall SPOT program, the validity of indicators used in the program, consistency across measurement, field conditions, training issues, scientific foundations of the program and/or behavioral detection methodologies, etc. In order to appropriately scientifically evaluate a program like SPOT, all of these and more would be needed.

How to Move Forward: Some Recommendations

• *Create a reliable research base of studies examining many of the issues related to security and the detection of deception*. Peer review, where and when possible, is particularly important. Shining a light on the process by making information on methodologies and results as open as possible (such as with devices like the polygraph, PCASS, voice-stress analysis, and neuroimaging) is necessary for determining if these technologies and devices are performing in a known and reliable manner. Clearly establishing the scientific validity of underlying premises, foundations, primitives, is essential. The larger the base of comparable scientific studies, the easier it is to establish the validity of techniques and approaches. A good example of this is the Bhatt and Brandon (2009) meta-analysis of the outcomes of studies in the literature related to voice stress analysis technologies. Similarly, the NRC *Threatening Communications* paper collection (2011) is an initial small step at establishing a body of literature on scientific approaches to understanding threatening communications and behavior.

• *Develop model systems, simulations, etc*. The use of model organisms in biology, such as *Drosophila* (a small fly) for helping to understand genetics and development, and *Aplysia* (the sea slug), for understanding neurons and memory, has spurred considerable scientific progress in these areas. Different kinds of model systems are needed for understanding behavior at the level of issues such as deception. Here we should look to the law enforcement community, the criminal justice system, and possibly border security, for models, approaches, analogies, data, and scientific guidance. Examples of

advances related to the complexity of behavior include well-known work on eyewitness identification (Loftus, 1996; Wells & Quinlivan, 2009).

• *Incorporate knowledge on the complexity, subtleties and idiosyncracies of human behavior*. Progress has been made on understanding how cognitive influences (Heuer, 1996; Pohl, 2004), psychological biases, and language use affect judgment, decision making, and risk assessment (Kahneman & Tversky, 1972; Thompson, 1999; Barrett, 2007). Also consider cultural and social contexts (Nisbett, 2003; Gordon, et al., in press).

• *Understand the interplay and differences between affect, emotion, stress, and other factors*. We have a tendency to oversimplify, categorize, and label complex behavior. The issues related to such matters can be seen in the contentious scientific debates on emotion and deception, discussed by other participants in today's hearing and summarized in part in a *Nature* article by Sharon Weinberger (2010). (See, also: Aviezer, et al., 2008; Barrett, 2006; Barrett, et al., 2007; Ekman, 1972; Ekman & Friesen, 1978; Ekman & O'Sullivan, 1991; Ekman, et al., 1999; Ekman, 2009; Hartwig, et al., 2006; Russell, et al., 2003; Widen, et al., in press.)

• *Make sure that we are not distracted or misled by the tools and toys that fascinate us*. While technological developments often hold considerable promise, they can be seductive and sometimes even can be counterproductive. The desire for automaticity and scale, coupled with urgent exigencies, should not reduce our need to attend to human aspects of the process and to the importance of devoting sufficient time to adequately understand behavior and manage interpersonal interactions.

• *Pay serious attention to the ethical issues and regulations related to human subjects research*, including 45 CFR 46 ("The Common Rule"), where applicable. Emerging areas include neuroethics (Farah, 2010) and autonomous agents (Wallach and Allen, 2010).

• *Reduce conflicts of interest* to the extent possible, particularly financial conflict of interest. The opportunity to profit from new and emerging technologies that have not been carefully and clearly scientifically validated and/or field evaluated, if necessary and possible, potentially puts our citizens, soldiers, and intelligence community at risk and could undermine our national security. We should have a clear understanding of both the strengths and weaknesses of tools, techniques, and technologies that are either being deployed or considered for future use.

• Develop an understanding of *how urgency, organizational structure, and institutional barriers can shape program development and assessment*. A detailed discussion of these issues is provided in the NRC *Field Evaluation* Workshop Summary (2010), summarized above in the *Field Evaluation* section. We should also strive to avoid the tendency to view results of the latest study as instantly confirming or falsifying controversial, new, or untested technologies (Mayew & Venkatachalam, in press). Consistency across multiple studies is essential.

• Support the importance of and need for *independent evaluation of new and controversial projects* and issues with appropriate scientific, technical, statistical, and methodological expertise. The NRC *Polygraph and Lie Detection* report (2003) provides a good case study for the importance of this point and the preceding bullet. Other

examples of such independent evaluations include many of the NRC reports listed in the *References* section, below. Another possible example is the JASON report on the SPOT program. Such reports should be seen as part of an iterative process that requires periodic modification and updating.

In our desire to protect our citizens from those who intend to harm us, we must make sure that our own behavior is not unnecessarily shaped by things like fear, urgency, institutional incentives or pressures, financial considerations, career and personal goals, the selling of snake oil, etc., that lead to the adoption of approaches that have not been sufficiently and appropriately scientifically vetted. To do so might ultimately end up being costly and counterproductive. We must not be distracted from the need for careful, well-considered, and well-established approaches for evaluating programs and technologies. We must be careful and thoughtful before investing in speculative or premature technologies that may be used out of desperation or because of potential commercial benefit. Where and when new technologies appear to be promising, we should obtain truly independent scientific expertise and assistance to provide context and guidance for the development possibilities and, if needed, for the consideration of appropriate metrics and methodologies for assessment and use. We should also keep in mind human costs and unintended consequences. As we all know, freedom and privacy must be considered in the context of safety and security. These values and goals are not incompatible. Sacrificing freedom and privacy to purchase illusory safety and security benefits only those who hope to harm us.

Chairman Broun, Ranking Member Edwards, and members of the Committee, I appreciate the opportunity to testify today. I would be happy to answer any questions that you might have about my testimony or related issues. Thank you.

REFERENCES

Aviezer, Hillel, Hassin, Ran R., Ryan, Jennifer, Grady, Cheryl, Susskind, Josh, Anderson, Adam, Moscovitch, Morris, and Bentin, Shlomo. (2008). Angry, disgusted or afraid? Studies on the malleability of emotion perception. *Psychological Science*, Vol. 19, No. 7, 724-732.

Barrett, Lisa Feldman. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, Vol. 1, #1, 28-58.

Barrett, Lisa Feldman, Lindquist, Kristen A., and Gendron, Maria. (2007). Language as context for the perception of emotion. *TRENDS in Cognitive Sciences*, Vol. 11, No. 8, 327-332.

Bhatt, S., and Brandon, S. E (2009). Review of voice stress-based technologies for the detection of deception. Unpublished manuscript, Washington, DC.

Chung, Cindy K. and Pennebaker, James W. (2011). Using computerized text analysis to assess threatening communications and behavior. In National Research Council, *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*. National Academies Press, Washington, DC, 3-32.

Damphouse, Kelly R. (2011). Voice Stress Analysis: Only 15 percent of lies about drug use detected in field test. *National Institutes of Justice (NIJ) Journal*, 259, 8-12.

Ekman, Paul. (1972). Universals and Cultural Differences in Facial Expressions of Emotions. In J. Cole (ed.), Nebraska Symposium on Motivation, 1971, University of Nebraska Press, Lincoln, Nebraska, 1972, 207-283.

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

Ekman, Paul. (2009). Lie catching and micro expressions. In Clancy Martin (ed.), *The Philosophy of Deception*. Oxford University Press.

Ekman, Paul and O'Sullivan, Maureen. (1991). Who can catch a liar? *American Psychologist,* 46(9), Sep. 1991, 913-920.

Ekman, Paul, O'Sullivan, Maureen, and Frank, Mark G. (1999). A few can catch a liar. *Psychological Science*, 10(3), May 1999, 263-266.

Farah, Martha J. (ed.). (2010). *Neuroethics: An introduction with readings*. The MIT Press, Cambridge, MA.

Gazzaniga, Michael S. (2011). Neuroscience in the courtroom. *Scientific American*, April 2011, 54-59.

Gordon, J. B., Levine, R. J., Mazure, C. M., Rubin, P. E., Schaller, B. R., and Young, J. L. (in press). Social contexts influence ethical considerations of research. *American Journal of Bioethic*s, 2011.

Hartwig, Maria, Granhag, Pär Anders, Strömwall, Leif A., and Kronkvist, Ola. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and Human Behavior*, 30(5), 603-619.

Heuer, Richards J., Jr. (1999). *Psychology of intelligence analysis*. Center for the Study of Intelligence, Central Intelligence Agency, Washington, DC.

Intelligence Science Board. (2006). *Educing Information: Interrogation: Science and Art*. The National Defense Intelligence College.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.

Loftus, Elizabeth F. (1996). *Eyewitness Testimony*. Harvard University Press, Cambridge, MA.

Mayew, William J. and Venkatachalam, Mohan. (in press). The power of voice: Managerial affective states and future firm performance. *Journal of Finance*, forthcoming.

Meloy, J. Reid. (2011). Approaching and attacking public figures: A contemporary analysis of communications and behavior. In National Research Council, *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*. National Academies Press, Washington, DC, 75-101.

Moreno, Jonathan D. (2006). *Mind Wars: Brain Research and National Defense*. The Dana Foundation, New York and Washington, DC.

O'Hair, H. Dan, Bernard, Daniel Rex, and Roper, Randy R. (2011). Communications-based research related to threats and ensuing behavior. In National Research Council, *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*. National Academies Press, Washington, DC, 33-74.

National Research Council. (2003). *The Polygraph and Lie Detection*. Committee to Review the Scientific Evidence on the Polygraph. Board on Behavioral, Cognitive, and Sensory Sciences and Committee on National Statistics, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC.

National Research Council. (2008). *Behavioral Modeling and Simulation: From Individuals to Societies*. Committee on Organizational Modeling: From Individuals to Societies. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC.

National Research Council. (2008). *Emerging Cognitive Neuroscience and Related Technologies*. Committee on Military and Intelligence Methodology for Emergent Neurophysiological and Cognitive/Neural Science Research in the Next Two Decades. Standing Committee for Technology Insight – Gauge, Evaluate, and Review Division on Engineering and Physical Sciences. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC.

National Research Council. (2008). *Human Behavior in Military Contexts*. Committee on Opportunities in Basic Research in the Behavioral and Social Sciences for the U.S. Military. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. Washington, National Academies Press, Washington, DC.

National Research Council. (2008). *Protecting Individual Privacy in the Struggle Against Terrorists*. Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals; Committee on Law and Justice (DBASSE); Committee on National Statistics (DBASSE); Computer Science and Telecommunications Board (DEPS). National Academies Press, Washington, DC.

National Research Council. (2010). *Field Evaluation in the Intelligence and Counterintelligence Context*. Workshop Summary. Planning Committee on Field Evaluation of Behavioral and Cognitive Sciences-Based Methods and Tools for Intelligence and Counterintelligence. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC.

National Research Council. (2011). *Intelligence Analysis: Behavioral and Social Scientific Foundations*. Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC.

National Research Council. (2011). *Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences*. Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC.

National Research Council. (2011). *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC.

National Science and Technology Council, Subcommittee on Social, Behavioral and Economic Sciences. Executive Office of the President of the United States. (2009). *Social, Behavioral and Economic Research in the Federal Context*. January 2009.

Nisbett, Richard E. (2003). *The Geography of Thought: How Asians and Westerners Think Differently... And Why*. Free Press.

Pohl, Rüdiger F. (2004). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*, Psychology Press, Hove, UK, 215–234.

Rubin, P. (2003). "Introduction." In S. L. Cutter, D. B. Richardson, & T. J. Wilbanks (Eds.), *The Geographical Dimensions of Terrorism*. Routledge, New York.

Rubin, P. and Wanchisen, B. (2011). "Introduction." In National Research Council, *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*. National Academies Press, Washington, DC.

Russell, James A., Bachorowski, Jo-Anne, and Fernández-Dols, José-Miguel. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54, 329-349.

Thompson, Suzanne C. (1999). Illusions of control: How we overestimate our personal influence. *Current Directions in Psychological Science*, 8(6), 187–190.

United States Department of Health and Human Services (HHS). (2009). Code of Federal Regulations. Human Subjects Research (45 CFR 46). (See: http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html)

United States Government Accountability Office (GAO). (2010). Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges. GAO-10-763, May 2010, Washington, DC.

Wallach, Wendell and Allen, Colin. (2010). *Moral Machines: Teaching robots right from wrong*. Oxford University Press, New York.

Weinberger, Sharon. (2010). Airport security: Intent to deceive? *Nature*, 465, 412-415.

Wells, Gary L. & Quinlivan, Deah S. (2009). Suggestive Eyewitness Identification Procedures and the Supreme Court's Reliability Test in Light of Eyewitness Science: 30 Years Later. *Law & Human Behavior*, 33, 1-24.

Widen, S. C., Christy, A. M., Hewett, K., and Russell, J. A. (in press). Do proposed facial expressions of contempt, shame, embarrassment, and compassion communicate the predicted emotion? *Cognition & Emotion*, in press, 1-9.